

GenCore version 5.1.6
Copyright (c) 1993 - 2004, Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:11:40 ; Search time 44 Seconds
(without alignments)
1016.836 Million cell updates/sec

Title: US-09-942-146A-1
Perfect score: 2451
Sequence: 1 MGRKEMVRDVPKMFVLISI.....PEHRAVNLSNSTSLWMLQ 465

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:*

1: piri:*

2: piri2:*

3: piri3:*

4: piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2416.5	98.6	466	2 S09837	hypothetical prote
2	122	5.0	475	2 G72274	glutaryl tRNA-Gln
3	116.5	4.8	1802	2 S69703	HKRI protein precu
4	111	4.5	1203	2 T31029	related to pathway
5	110.5	4.5	739	2 T21769	hypothetical prote
6	109	4.4	651	2 C56653	membrane glycoprot
7	108.5	4.4	1224	2 E71611	hypothetical prote
8	107.5	4.4	919	2 T16693	hypothetical prote
9	106	4.3	1285	2 S70582	botulinum neurotox
10	103.5	4.2	2101	2 S7245	insulin receptor (
11	103	4.2	2028	2 T08025	DNA-directed RNA p
12	102.5	4.2	505	2 B46629	mucin 6, gastric (
13	102.5	4.2	638	2 T38863	hypothetical prote
14	101.5	4.1	1169	2 S38181	flocculation prote
15	101.5	4.1	1848	2 A44140	cellulose-binding
16	101	4.1	880	2 S60137	beta-N-acetylhexos
17	100.5	4.1	1995	2 T08166	probable membrane
18	100.5	4.1	3570	2 T45035	mucin MUC5B, trach
19	100	4.1	660	2 JW0067	chitinase (EC 3.2.
20	100	4.1	1429	2 T41699	C2-domain family p
21	99.5	4.1	369	2 C88030	protein F46P5.10 (
22	99.5	4.1	1348	2 AH115	cell surface prote
23	99	4.0	375	2 H70226	conserved hypotet
24	99	4.0	621	2 C81026	ABC transporter, A
25	98.5	4.0	738	2 T44194	hypothetical prote
26	98.5	4.0	738	2 T44007	hypothetical prote
27	98.5	4.0	1365	2 T18419	hypothetical prote
28	97.5	4.0	1174	1 HUBYDH	helicase (EC 3.6.1
29	97	4.0	452	2 B97095	membrane associate

30	97	4.0	621	2 H81971	probable ABC trans
31	97	4.0	1777	2 T34369	hypothetical prote
32	96.5	3.9	347	1 DNHUN2	NADH2 dehydrogenas
33	96.5	3.9	3890	2 C89921	hypothetical prote
34	96	3.9	721	2 B90487	hypothetical prote
35	96	3.9	722	2 T00049	hemocyte protein A
36	96	3.9	809	2 T39626	hypothetical prote
37	96	3.9	1234	2 B36186	I factor protein 2
38	96	3.9	2437	2 T18482	hypothetical prote
39	95.5	3.9	474	2 B75126	hypothetical prote
40	95.5	3.9	653	2 H82316	conserved hypotet
41	95	3.9	291	2 A99181	xerC/D integrase-r
42	95	3.9	522	2 T37536	cell polarity prot
43	95	3.9	534	2 T33393	hypothetical prote
44	95	3.9	796	2 T21460	hypothetical prote
45	95	3.9	899	2 C71608	origin recognition

ALIGNMENTS

RESULT 1

S09837

hypothetical protein UL74 precursor - human cytomegalovirus (strain AD169)
C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004

C:Accession: S09837

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.;

M.J. Barrell, B.G.

Cur. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis-of-the-protein-coding-content-of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09837

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <CHE>

A:Cross-references: UNIPROT:P16750; EMBL:X17403; NID:G59591; PIDN:CAA35389.1; PID:G17808

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus hypothetical protein UL74

C:Keywords: glycoprotein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-466/Product: hypothetical protein UL74 #status predicted <MAT>

F:75,83,87,103,130,157,162,171,219,242,288,292,350,367,385,392,399,433,443,454/Binding

Query Match 98.6%; Score 2416.5; DB 2; Length 466;

Best Local Similarity 98.9%; Pred. No. 4.5e-175;

Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MGRKEMVRDVPKMFVLISISFLLVSPFINKVMSKALYNRPWGLVLSKIKYKLDQLKL 60

Db 1 MGRKEMVRDVPKMFVLISISFLLVSPFINKVMSKALYNRPWGLVLSKIKYKLDQLKL 60

Qy 61 EILRQLETTISTKYNVSKQPVKNLTMTAEPPOYVILAGPIQNTSYITVLFNDFYSTQLRK 119

Db 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEPPOYVILAGPIQNTSYITVLFNDFYSTQLRK 120

Qy 120 PAKVYVSOYVHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179

Db 121 PAKVYVSOYVHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

Qy 180 NVPRWNTKLYVGPVKVNVDSQTIYFLGLTALLLYAQRNCTHSPFLVNMARNLFRVPKY 239

Db 181 NVPRWNTKLYVGPVKVNVDSQTIYFLGLTALLLYAQRNCTHSPFLVNMARNLFRVPKY 240

Qy 240 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPYFSYTTSAALNVTNTVYSITTA 299

Db 241 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPYFSYTTSAALNVTNTVYSITTA 300

Qy 300 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVVTTLRYRONPCEPSRNTAVSEFMKN 359

Db 301 ARRVSTSTIAYRPDSSFMKSIATQLRDLATVVTTLRYRONPCEPSRNTAVSEFMKN 360

Qy	360	THVLRNETPTIYGTILDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD	419
Db	361	THVLRNETPTIYGTILDMSSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD	420
Qy	420	YLDSELLFDEIRNFSLSRSPPTVNLTPPEHRAVNLSTNSLWMLQ	465
Db	421	YLDSELLFDEIRNFSLSRSPPTVNLTPPEHRAVNLSTNSLWMLQ	466
RESULT 2			
glutamyl tRNA-Gln amidotransferase, subunit A - Thermotoga maritima (strain MSB8)			
C:Species: Thermotoga maritima			
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004			
C:Accession: G72274			
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey			
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.			
C.M.			
Nature 399, 323-329, 1999			
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq			
A:Reference number: A72200; MUID:99287316; PMID:10360571			
A:Accession: G72274			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-475 <ARN>			
A:Cross-references: UNIPROT:Q9X0Z9; GB:AE001782; GB:AE000512; NID:94981821; PIDN:AAD3634			
A:Experimental source: strain MSB8			
C:Genetics:			
A:Gene: TM1272			
C:Superfamily: indoleacetamide hydrolase			
Query Match 5.0%; Score 122; DB 2; Length 475;			
Best Local Similarity 21.2%; Pred. No. 0.16;			
Matches 101; Conservative 64; Mismatches 171; Indels 140; Gaps 23;			
Qy	17	LISISFLYSFINCKVMSKALYNRPWGLVLKIGIKYKLDQLKRLRQLE----	TTIST 72
Db	1	MIDLDFRKUTIEC-----LKLSEEREKLPQLSLETIKELDPHKAFISV	46
Qy	73	KNVS-----KQPVKNLTMTTEFPQYILAGPTQNSITLWDFYSTQLR	118
Db	47	RENVSEKKGKFWGIPVAIKONILTLGMRIT-----CASRIENYESV---	FDAIVVKKM 98
Qy	119	KPAKYV-----YSQYNHTAKITP---RPPCGRVP-----SMTCLSE-MLNVSK	159
Db	99	KEAGFVVVGKANLDEFAMGSSSTERSAFFFPTRNPDWLERVPGSSGSAASVAGMVVAAL	158
Qy	160	RNDTG-----EQCGNFTTFNPMFFNVPWNTKLY-----VGP-TKVNVDSTIYFLG	206
Db	159	GSDTGGSVRQPSILCG-VVGKFTYGLVSRYGLVAFASSLDQIGPITKTVRDA-----	210
Qy	207	LTALLRAYAQRNCTHSYLVNMSRLNRPVKYINGTKLKNTRKLRK-QAPVKEQPEK	265
Db	211	--AIIIMEIISORDNATTNRRKVDFLSEIEEGVSGMKFAVPBEIYEHDIIEGVSRFEE	268
Qy	266	KAK-----KTQSTTPPYFSYTTSAALNVTNTVYSITTAARVSTSTIARPDSSFMK	318
Db	269	ALKLELRLGAKVERKLPKIKYSA-----TYVVIAPAE--ASSNLARPDGVKYL	317
Qy	319	SIMATQLRLDATWYTTLRYQRNPFCEPSRNRNRAVSEFMKNVHLRNPTIYIGTLDM	378
Db	318	RIKEKGLREM-----YMKTRNVGFGSEVRRIMI-----GTFTL 351	
Qy	379	SSLY-----NETMFVENKTASDSNKT-----TPTSPSMGQRTFI-DPLWDYL	421
Db	352	SAAYEAYFNKAMKVRKRKISDELNEVLSDYDAULTPTSPYAFKIGIKDPLTYVL	407
RESULT 3			
HKR1 protein precursor - yeast (Saccharomyces cerevisiae)			
N:Alternate names: protein YDR420W			
C:Species: Saccharomyces cerevisiae			
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004			
C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 09-Jul-2004			
C:Accession: S69703; A53382			
R:Dietrich, F.S.			
A:Description: The sequence of S. cerevisiae lambda 3641 and cosmids 9461, 9831, and 941			
A:Reference number: S69555			
A:Accession: S69703			
A:Molecule type: DNA			
A:Residues: 1-1802 <DIE>			
A:Cross-references: UNIPROT:Q04051; EMBL:U33007; NID:9327685; PIDN:AAB64857.1; PID:99276			
R:Kasahara, S.; Yamada, H.; Mio, T.; Shiratori, Y.; Miyamoto, C.; Yabe, T.; Nakajima, T.			
J. Bacteriol. 176, 1488-1499, 1994			
A:Title: Cloning of the Saccharomyces cerevisiae gene whose overexpression overcomes the			
A:Reference number: A53382; MUID:94156857; PMID:8113191			
A:Accession: A53382			
A:Molecule type: DNA			
A:Residues: 1-581, 'A', 593-593, 'A', 595-1802 <KAS>			
A:Cross-references: EMBL:S69101; NID:9545659; PIDN:AAB30051.1; PID:9545660			
A:Experimental source: YNN295			
A>Note: sequence extracted from NCBI backbone (NCBIN:144410, NCBI:P:144411)			
C:Genetics:			
A:Gene: SGD:HKR1			
A:Cross-references: SGD:S0002828; MIPS:YDR420W			
A:Map position: 4R			
C:Keywords: calcium binding; glycoprotein; transmembrane protein			
P:1-21/Domain: signal sequence #status predicted <SIG>			
P:22-1802/Product: HKR1 protein #status predicted <MAT>			
P:1483-1508/Domain: transmembrane #status predicted <TM>			
P:1645-1656/Domain: calcium binding #status predicted <CAL>			
Query Match 4.8%; Score 116.5; DB 2; Length 1802;			
Best Local Similarity 20.7%; Pred. No. 2.5; Indels 127; Gaps 19;			
Matches 90; Conservative 60; Mismatches 157; Indels 127; Gaps 19;			
Qy	68	TTISTKYNVSKQPVKNLTMTTEFPQYILAGPTQNSITLWDFYSTQLRPAKVYISQ	127
Db	783	TSITSPYDIVYSESTAAISSGYT-----PSPSAVA-----MSSTSSSPYDIVYSL	830
Qy	128	YNHTAK-----TITFRPPCCGRVPSMTCLSEMLNVKRNDTGEOCGNFTTENPME---	FN 180
Db	831	SSGASRSSIATYFSPSPSTSLPTSS-----TYTFSSAYAFES 870	
Qy	181	VPRNWKLYGPTKVVVD-SQTIYFLGLTALLR--YAQRNCTHS-----FYLVN 227	
Db	971	SERYSTSTTAPTQIHSTLSRIYDFLLQTSMAIQSVSQISTSTLNDEIHSSALSVFN	930
Qy	228	AMSNLFRVPKYINGTKLKNTRKLRKQAPVKEQPEKAKKT-----QSTTT 275	
Db	931	PSASNLVETSLIISSTQASITSPKNSAKISSLSQSLSSSTKNPYDTANKNTETSGRSTVV	990
Qy	276	PYFSYTTSAALNVTNTVYSIT-TAARRVSTSIAY-----RP 312	
Db	991	SNFLYTSAA--KPDNEKFSATPTEITTISSSHAYSLSPSSHNSVTGLSHNFPVDSKS	1048
Qy	313	DSSF-----MKSIMATQLRLDATWYTTLRYQRNPFCEPSRNRNRAV	353
Db	1049	ATSGYSSSSISSIKLSKETIPASKSVSNITQER--ITSFTSTL--RANSQSEKSEGRNSV	1104
Qy	354	SEFMKNVHLRNPTIYIGT-LDMSLY-----NETMFVENKTASDSNKTTP	402
Db	1105	GS-LQSSHI-----SSNPSTLNTKVDKSLSRKVKTMGNGEETGLTTTKIQYKSSSETS	1160
Qy	403	TSPSMGQRTFIDP 416	
Db	1161	GSYSRSFTKISIGP 1174	
RESULT 4			
TS1029			
related to pathway-specific nitrogen regulator [imported] - Neurospora crassa			
N:Alternate names: protein B7F21.100			
C:Species: Neurospora crassa			
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004			

A;Gene: C8SP:F35BL2.7
A;Map position: 5
A;Introns: 26/1; 61/3; 92/2; 131/1; 164/3; 240/1; 295/3; 319/2; 428/2; 482/3; 704/3

Query Match 4.5%; Score 110.5; DB 2; Length 739;
Best Local Similarity 21.3%; Pred.No.2.1;
Matches 83; Conservative 42; Mismatches 151; Indels 113; Gaps 17;

Qy 79 QPV-----KNLTMTNTEPQQYILAGP-----IQNYSTIYL-----WPDF-----Y 113
Db 313 QPVLYKYPYSTNSNSSFPQ--LIPSPMATPHIHNASVTFVVVKPGVPQNDFAAAPQSRSVY 370

Qy 114 STQLRKPA-----KYIVSQ-----YNHTAKTITFR-----PPGGRVPSMTCLEMLN 156
Db 371 SPQIWNPAAPEKDFTFSDPKVYNFSINLQTVKLENDELDGVQSVNDGLTLTKKYN 430

Qy 157 VSKRNTDGECCGFNTFNPMFFNVPRWNKL-----YYGP TKANVDSTIIFLGITAL 210
Db 431 KDKSADAYLSGIQR-----YLNLYKTSGTSSEVINFMIDRLMPDGTT-----474

Qy 211 LLRYAORNCTHSPYLVNAMSRNFLRPVPKYINGTKLXNTMKLRKOAPVKEOFEKKAKKT 270
Db 475 -----QSPGITTVVPSGSFST-----APIVTSVSTRATRV 503

Qy 271 QSITTPYSVTSSAALN---VTNN-----VTVSIITAARVSSTAIYRP---DSFPMKS 319
Db 504 SSSATPVSVTVTITSQSIIVTNKSSPITSKTITTSINVSSSTKTRPMQSTTTYQKL 563

Qy 320 IMATQLDLATWYTILRYQRNQPCPSRNRTAVSFEMKNTHVLIRNETPTTIYGILDMS 379
Db 564 AQSASTASTPIBITTONISTASIAPS-TKTPTPLSLITRSSTTKLSDIVI----TKS 618

Qy 380 SLYYNETMFVENKTASDNKTTPTTSPSMG 408
Db 619 SLAYTATPTSAA PSTSALTITHTTPIP 647

RESULT 6
C56653
membrane glycoprotein precursor - human herpesvirus 6 (strain UI102)
C:Species: human herpesvirus 6
C>Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 09-Jul-2004
C:Accession: U56653
R;Gompels, U.A.; Carss, A.L.; Sun, N.; Arrand, J.R.
DNA Seq. 3, 25-39, 1992
A>Title: Infectivity determinants encoded in a conserved gene block of human herpesvirus 6
A:Reference number: A56653; UID:93091236; PMID:1333836
A:Accession: C56653
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <GOM>
A:Cross-references: UNIPROT:Q06093; GB:X83413; NID:g853961; PIDN:CAAS8381.1; PID:g8540202
A:Experimental source: strain UI102
A>Note: sequence extracted from NCBI backbone (NCBIN:120533, NCBI:P:120546)
C:Superfamily: herpesvirus glycoprotein H
C:Keywords: glycoprotein

Query Match 4.4%; Score 109; DB 2; Length 651;
Best Local Similarity 18.3%; Pred.No.2.3;
Matches 107; Conservative 77; Mismatches 189; Indels 192; Gaps 25;

Qy 5 EMVRDPVMFVLISIFSLLVSPINCKWSKALYNRPWRGLVLSKIG-----51
Db 4 EVIVQSYKSKSYFSHTFYLYKEI-----VVNSP-DMLHISRGLFLGLFAIVMHVS 54

Qy 52 ----KYKLDQLKBELRQLETISTKYNVSKQPVNLTMTNEFPQQYILAGPINQYSITYL 108
Db 55 NLIIKTYSDPLE-----AFKVNRHNWSDEQREHF-----YDLRNL 89

Qy 109 WFDIFYSTOLRRKPAKYVYSQNHAKTITFRPPPGRVPSMTCLSEMNLVSKRNDTGEOGC 168
Db 90 YTSFCQTNL-----SLDCFTQILTNTVFNSWDIRDSC 120

169 GNFTFPMFNVPRWNTKLYVGPTKVN---VDSQTIYFL----- 205
 121 KSAVLSPL-QNLPRTEIKVLSSTANKSIIASSFSLYLLPATLTYTADPPCVLLP 179
 206 -----GLTALLLYAQRNCHTSFYLVNMSRNLFRVPKYNGTKLQNTM 249
 180 FKILGAQLFDIKLTEESLRMSKFSNLTSLRS--LTSFTSKNFFNYTSFY-FLLYNT- 235
 250 RKLKRKAQVKEQEKAKK---TQS-----TTTTPYFSYTSAAAL----- 286
 236 ----TSCVPSNDQYFKQSPKPIVNTTSGRAIVNFDLSILITTP---SSTSASLTSPHPS 288
 287 -NVNTNVTYSITTAARVSTSIAYRDPSSPMKIMATQLR-----DLATWVYT----- 334
 289 TNITPAPPVTKNSKLTHTDIKVNTPTPIITQTIESIKKIVKRSDFPRPMTPTDIP 348
 335 TLRVQRNPFCSRRNRATAVSEFMKNTHVLINNETPYIYGLDMSSLYNNTMVEKNKT- 393
 349 TLTLRLNATIKTEQNTENPKSPKPTNF--ENTTIRIPKTLGATATNATQKIESWTF 405
 394 ----ASDSNKTPTSP-----SMGQRT--FID-----PLWDYLDLSLLFLDIRNPSL 435
 406 TTIGIKINGNTYSPKSNYILKSKSQSTTKFTDAEHTTPI---LKFTTWQNTVRTYMS 462
 436 RSPTVNLTPPEHRAVNLSTSNL 460
 463 HNTVQNWTDKFOR--TTLKSSNEL 485

RESULT 7
 E71611
 hypothetical protein PFB0580w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
 C:Accession: E71611
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ;; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: E71611
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1224 <GAR>
 A:Cross-references: UNIPROT:O96209; GB:AE001404; GB:AE001362; NID:g3845219; PIDN:AAC7190
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0580w

Query Match 4.4%; Score 108.5; DB 2; Length 1224;
 Best Local Similarity 21.7%; Pred. No. 6;
 Matches 78; Conservative 47; Mismatches 116; Indels 119; Gaps 19;

73 KYNSKQPVKNLTWNTE--FPQYI-LAGPQYISITYLW----FDP----- 112
 82 KYRNNNNVKNILKDEILDYXNLHLSNLHNDINFIYSSNNIFLNNKNNKPKYPS 141
 113 -YSTQLRKPAK-----YVYSQNHAKTITTRPPCPGRVPSMTCLSEMLN-----VSKRN 161
 142 KNSNEIKKHKNVKNVYNTNHHYTKKKNFYSNPT-EVNTNSLLSNLKNLSLYSPRK 200
 162 DTGEGCGNFTTPEMPFNVPRWNTKLYVGPTKYNVDSTQIYFLGLTALLRYAQRNCH 221
 201 DT-----SNFN---PSCDKNNT-----TFSPKNCLH 223
 222 SFYLVNAMSR-----NLFRVPKYNGTKLQNTMRKLR-KOAPVKEQEKAKKTQSTTT 275
 224 E---SNPSTSTCPVNVNTIPLAIN--LLNNVNDDISPIHPLPSESSTASSTASAST 278
 276 PYFSYTSAAALNVTNVTYSITTAARVSTSTIAYRDPSSPMKIMATQLRDLATWVYT 335
 279 ---SASTSASTSVSTSVSTSA---STTMNSPRPSDNHISNPSPLSRESRATEQVNR 332

336 LRY-----RQNPFCBSPSRNRRTAVSEPMKNTHTVLIRNETPYTI 372
 333 LYFPVNDVTSKSDPNPNNELTNNMPKHPICEBTRN-----DNHG--IRNSIYPL 382

RESULT 8
 T16693
 hypothetical protein R05F9.12 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T16693; T27894
 R:Hallsworth, K.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid R05F9.
 A:Reference number: Z18559
 A:Accession: T16693
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-919 <HAL>
 A:Cross-references: UNIPROT:Q21750; EMBL:U41533; NID:g1109807; PID:g1109820; PIDN:AAA831
 R:Hallsworth, K.
 submitted to the EMBL Data Library, June 1995
 A:Description: The sequence of C. elegans cosmid ZK546.
 A:Reference number: Z20437
 A:Accession: T27894
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-575 <HA2>
 A:Cross-references: EMBL:U29380; PIDN:AAA68743.1; CESP:ZK546.10
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:R05F9.12; CESP:ZK546.10
 A:Introns: 31/3; 94/2; 174/3; 329/3; 524/3; 575/2; 607/2; 645/2; 739/1; 816/3
 C:Superfamily: lysosomal alpha-glucosidase; sucrose/isomaltase homology; trefoil homology

Query Match 4.4%; Score 107.5; DB 2; Length 919;
 Best Local Similarity 19.2%; Pred. No. 4.8;
 Matches 102; Conservative 67; Mismatches 210; Indels 153; Gaps 24;

58 LKLEILLQLETTISTKYNVSKQPVKNLTM-----NTEFPQYIILAGPQYISITY----- 107
 374 IEWETKAQVQTALQNLVPMKADTKIMLGVVWPDNHNVAFPDFLDSTNNTNWINEFVNYQ 433
 108 -----LWDF-----YSTQLRKPAKYVSYQVNHAKTITFRP-----PPCGRVPSMT-- 149
 434 SQVAFDGIWIDMNEPSNFGTNQDHP--WYFDSDDHPNDAPLFCPTNGSSPWEMPYKTRA 491
 150 -----CLSEMLNVSKR---NDTGEQCGNFTTFNP-----MFF 179
 492 VWRFGDANGAFLSSNTLCMLAQODGKQRFYNATGKGVAVSRSTYPSAGRYAGHWLGD 551
 180 NVPRWN--TKLYVGPTKVN-----VDSQTIYFLGLTA--LLRYAQRNCHTSFY----L 225
 552 NTARWEDLRTSVIGAQEFNLFIPYVGSVDVCGFTTTEELCLRWQMGAFSPFRNHT 611
 226 VNAMSRNLFRVPKYNGTKLQNTMRKLRKOAPVKEQEKAKKTQSTT--TPYFSYTS 283
 612 IGAPAQDPAPWPSVAATAKCANLFR---YQVLPVLSLHFTASLSGATVIRPVFEYPTD 668
 284 AALNVTTNVTYSITTAAR-----RVSTSTIAYRDPSSPMKIMATQLRDLATWVYT 335
 669 AE---TFNLGYFMWGSRIILVAPVIYQTTSVNAYLPTD-----RW-YSL 709
 336 LRYQRNPFCE-----PSRNRATAVSEPMKNTHTLIR-----NETPYTY----- 373
 710 FDYRYSIMSGYATVPAPTTSRIPVVRGYSVIPROTPTSTTTTATRSNPPELLIAPCPL 769
 374 ---GTL-----DMGSLYYNETMVFENKTASDSNKTTPS-----PSMGFOR 411
 770 GMGEGTLWDGGETIVNDFNSYDHFQDFMNVNATGCGSVTIITHSKSKSLSLTLDIE 829
 412 TFDPLWDYLDLSLLFLDIRNFSRSTYVNLTPPEH---RRAVNLSTSNL 460

Db 830 IFNPSAPNFRSFTINGKLVNPNVQBSYSGITKILYISTEGLVNLSSADSI 881

RESULT 9

botulinum neurotoxin type Daa precursor - Clostridium botulinum phase d-eA

C:Species: Clostridium botulinum phase d-eA

A:Note: host Clostridium botulinum type D (strain South Africa)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000

C:Accession: S70582

R:Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.

Biochim. Biophys. Acta 1307, 123-126, 1996

A:Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C and E

A:Reference number: S70582; MUID:96283801; PMID:8679691

A:Accession: S70582

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1285 <MOR>

A:Cross-references: EMBL:D38442; NID:gl374775; PIDN:BA07477.1; PID:gl374776

C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neuromuscular transmission by binding to the presynaptic membrane and forming a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membrane.

C:Superfamily: tetanus toxin

C:Keywords: disulfide bond; neurotoxin; transmembrane protein

F:1-447/Product: botulinum neurotoxin type Daa light chain #status predicted <MARI>

F:448-1285/Product: botulinum neurotoxin type Daa heavy chain #status predicted <MAT>

Query Match 4.3%; Score 106; DB 2; Length 1285;

Best Local Similarity 22.0%; Pred. No. 9.8;

Matches 85; Conservative 52; Mismatches 132; Indels 118; Gaps 22;

Qy 87 NTEPPQYILAGPTQNYISITVLPDFYSTQLRKPA-----KVYQYNHTAKTIIFRPPPC 142

Db 264 NVQPEELYTFGG-----SDVEIIPQIERLQRLREKALGHYKDIKRLNNINKTI----- 311

Qy 143 GRVPS-----MTCLSEMLNVSQRNDTGEQCGNFTFNPFPNPNVNTKLYVGPT 193

Db 312 -----PSSWSSNDKYKKIFSEKYNFDKN-----TGNP-----LVNIDKFS-LYSDLT 355

Qy 194 KVNVDSTQYIFGLTALLRYAQRNCTHSP---YL-----VNAMSNRLFRVPKYIN-----GT 243

Db 356 --NVMSVYI-----SSQNVNKRTHYFSKHYLPVFNILDDNIYTIINGFNLTGKF 406

Qy 244 KLNMTMKLRKQAPVKEQFEKAKKTOSTTTPYFSTSAALNVTNTVYISITTAAR-- 301

Db 407 NIENSGQNIERNPALQK-----LSSESVDLFTKVCRLRLTRNSRDD 447

Qy 302 ----RVSTSTIAYRPD--SSFMSKSIATOLRLATVYTLTLYRONPFCPS-----RN 349

Db 448 STCIQVKNLTPYVADKDSISQEIFESQIITDETNE---NYSNFSLDESILDAKVPTN 504

Qy 350 RTAVSEFMKNTHVLRNPTTYIG-----TLD-----MSSLYYNETMFVENKTASDS 397

Db 505 PEAVDPLLPNV-----NMEPLNVGEEBEVFYDDITKQVDYLNYSYYLEAQKLSN-NVENI 558

Qy 398 NKTPTSPSGMGRFTFDPLWDYLDL 424

Db 559 TLTTSVVEALGYS---NKIYTLFPLSL 581

RESULT 10

S57245

insulin receptor (version 2) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Jun-2000

C:Accession: S57245

R:Fernandez, R.; Tabarini, D.; Azpiroz, N.; Frasc, M.; Schlessinger, J.

EMBO J. 14, 3373-3384, 1995

A:Title: The Drosophila insulin receptor homolog: a gene essential for embryonic development

A:Reference number: S57245; MUID:95354655; PMID:7628438

A:Accession: S57245

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2101 <PER>

Db 830 IFNPSAPNFRSFTINGKLVNPNVQBSYSGITKILYISTEGLVNLSSADSI 881

RESULT 9

botulinum neurotoxin type Daa precursor - Clostridium botulinum phase d-eA

C:Species: Clostridium botulinum phase d-eA

A:Note: host Clostridium botulinum type D (strain South Africa)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000

C:Accession: S70582

R:Moriishi, K.; Koura, M.; Abe, N.; Fujii, N.; Fujinaga, Y.; Inoue, K.; Ogumad, K.

Biochim. Biophys. Acta 1307, 123-126, 1996

A:Title: Mosaic structures of neurotoxins produced from Clostridium botulinum types C and E

A:Reference number: S70582; MUID:96283801; PMID:8679691

A:Accession: S70582

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1285 <MOR>

A:Cross-references: EMBL:D38442; NID:gl374775; PIDN:BA07477.1; PID:gl374776

C:Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neuromuscular transmission by binding to the presynaptic membrane and forming a disulfide bond. The heavy chain mediates the binding of toxin to the presynaptic membrane.

C:Superfamily: tetanus toxin

C:Keywords: disulfide bond; neurotoxin; transmembrane protein

F:1-447/Product: botulinum neurotoxin type Daa light chain #status predicted <MARI>

F:448-1285/Product: botulinum neurotoxin type Daa heavy chain #status predicted <MAT>

Query Match 4.3%; Score 106; DB 2; Length 1285;

Best Local Similarity 22.0%; Pred. No. 9.8;

Matches 85; Conservative 52; Mismatches 132; Indels 118; Gaps 22;

Qy 87 NTEPPQYILAGPTQNYISITVLPDFYSTQLRKPA-----KVYQYNHTAKTIIFRPPPC 142

Db 264 NVQPEELYTFGG-----SDVEIIPQIERLQRLREKALGHYKDIKRLNNINKTI----- 311

Qy 143 GRVPS-----MTCLSEMLNVSQRNDTGEQCGNFTFNPFPNPNVNTKLYVGPT 193

Db 312 -----PSSWSSNDKYKKIFSEKYNFDKN-----TGNP-----LVNIDKFS-LYSDLT 355

Qy 194 KVNVDSTQYIFGLTALLRYAQRNCTHSP---YL-----VNAMSNRLFRVPKYIN-----GT 243

Db 356 --NVMSVYI-----SSQNVNKRTHYFSKHYLPVFNILDDNIYTIINGFNLTGKF 406

Qy 244 KLNMTMKLRKQAPVKEQFEKAKKTOSTTTPYFSTSAALNVTNTVYISITTAAR-- 301

Db 407 NIENSGQNIERNPALQK-----LSSESVDLFTKVCRLRLTRNSRDD 447

Qy 302 ----RVSTSTIAYRPD--SSFMSKSIATOLRLATVYTLTLYRONPFCPS-----RN 349

Db 448 STCIQVKNLTPYVADKDSISQEIFESQIITDETNE---NYSNFSLDESILDAKVPTN 504

Qy 350 RTAVSEFMKNTHVLRNPTTYIG-----TLD-----MSSLYYNETMFVENKTASDS 397

Db 505 PEAVDPLLPNV-----NMEPLNVGEEBEVFYDDITKQVDYLNYSYYLEAQKLSN-NVENI 558

Qy 398 NKTPTSPSGMGRFTFDPLWDYLDL 424

Db 559 TLTTSVVEALGYS---NKIYTLFPLSL 581

RESULT 10

S57245

insulin receptor (version 2) - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 02-Jun-2000

C:Accession: S57245

R:Fernandez, R.; Tabarini, D.; Azpiroz, N.; Frasc, M.; Schlessinger, J.

EMBO J. 14, 3373-3384, 1995

A:Title: The Drosophila insulin receptor homolog: a gene essential for embryonic development

A:Reference number: S57245; MUID:95354655; PMID:7628438

A:Accession: S57245

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2101 <PER>

A:Cross-references: EMBL:U28136

C:Genetics:

A:Gene: FlyBase:InR

A:Cross-references: FlyBase:FBgn0013984

C:Superfamily: Drosophila insulin receptor; protein kinase homology

C:Keywords: ATP; receptor

F:1321-1609/Domain: protein kinase homology <KIN>

F:1329-1337/Region: protein kinase ATP-binding motif

Query Match 4.2%; Score 103.5; DB 2; Length 2101;

Best Local Similarity 20.8%; Pred. No. 29;

Matches 99; Conservative 68; Mismatches 175; Indels 135; Gaps 23;

Qy 44 GLVLSKIGKYKLDQLKLEILRLQLETTISTKYNVSKQPVKNLTMT-----EFPYYILA 97

Db 508 GCVIDKNGN-----ESCISCR-NVS---FNNICMDSCKPGYQFDSRCVTA 549

Qy 98 GPIQNYSTVYLPDFYSTQLRKPAKYVYVQYNHTAKTIIFRP-----PPC-GR 144

Db 550 ----NECITLTKPTNS-----VSGIPYNGQCITHCPTGYQKSENKRMCEPCGG 596

Qy 145 VPSMTCLSEMLNVSQRNDTGEQCGNFTFNPFPNPNVNM-----TKLYVGPTKVNVD 199

Db 597 KCDKECSGLID-SLERAREFHGCTIITGTEPLTISIKRESGAHVMDLKYGLAAVHKIQ 655

Qy 200 QTI-----YFLGLTALLRYAQRNCTHSPFYNVAMSNRLFRVPK----- 238

Db 656 SSLMHLTYGLKSLKPFQSLTEISGDPMDADKALYVLDNRDLDELWGPNTVPIRKG 715

Qy 239 ---YINGTKLNTMRKLRKQAPVKEQFEKAKKTOST-----TTPYFSY 283

Db 716 VFFHFNKLCVSTINQLLPLMLASKPFKEKSGDEGADNGRSGCTAVNLVLOSVGANS 775

Qy 284 AALNVTNTVYISITTAARRVSTSTIAYRPDSSFMKSI-----ATOLRLATVYVYTL 336

Db 776 ASLNTTKVEIG---EPQKSNATIVKQPAFGFVYHMDVPGTOLK--AVTIHAMI 830

Qy 337 RYRONPFCPSRNNTAVSEFMKNTHVLRNPTTYIGTLDMSLYYNETMFVENK-TAS 395

Db 831 AGK-----VSSPEKSGVMVLSNLPIPTNYS-----YVVRTMAISSELTNA 870

Qy 396 DSN-----KTTPTSPS---MGFQRTFDPLWDYLDLFLDE--IRNFSLSPTVYN 442

Db 871 ESDVKNFRTPNPGRSKVVAITSDSKINVTWYSYLDNLMLALTRYFIKAKLINRPTNN 927

RESULT 11

T08025

DNA-directed RNA polymerase beta' chain homolog - Chlamydomonas reinhardtii chloroplast

N:Alternate names: rpoC2 protein

C:Species: Chloroplast Chlamydomonas reinhardtii

C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 09-Jul-2004

C:Accession: T08025

R:Nuotio, S.; Purton, S.

submitted to the EMBL Data Library, May 1996

A:Description: The chloroplast rpoC2 gene of Chlamydomonas reinhardtii.

A:Reference number: Z16298

A:Accession: T08025

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2028 <NUO>

A:Cross-references: UNIPROT:Q7PCJ6; EMBL:U57326; NID:gl354831; PIDN:AAB01997.1; PID:gl35

C:Genetics:

A:Gene: rpoC2

A:Genome: chloroplast

C:Keywords: chloroplast

Query Match 4.2%; Score 103; DB 2; Length 2028;

Best Local Similarity 20.8%; Pred. No. 31;

Matches 111; Conservative 83; Mismatches 204; Indels 136; Gaps 29;

Qy 13 KMFLVI-----SISFLVSVFINKVMSKAL---YNRPWRLGLVLSKIGKYLQDLKLELRQ 65

Db 678 KLLKLINKHSIALNOTKPIINTSLLKGHLVAYARP-VFIITNKAEPPIAKQKDLILLP 736
QY 66 LETTISTKYNSKQPVKNLTWNTFFP-----QYILAGP-----IONYSITYLWDFDYS 114
Db 737 KNAIINVL--IKPOEKNSLKAIFPIGGNAANNHSIFVSPNTKLANPNVSYLKHHIYF 794
QY 115 TQLR-----KPAKYVYSQVNHAKTITFRPPCGRVPSMTCLSE-----MLNYSKEN--- 161
Db 795 NQMPFFDPPFNASYLESYTENSIKPLT-----KVDPMHSFAQKHKLPESEKRL 847
QY 162 -----DTGQCGN--FTTNPMFNVPRWNTKLYVGPT---KVNVD-SQTIYFLGLTAL 210
Db 848 QFKALHPKQPCMLCFCNSNLIFNSKTTNSLVIRKTTNYKYNIDLSEGVDFKLLKTN 907
QY 211 LLRYAQRNCTHSFVLVNMARNLPR-VPKVIINGTKLNTMRKLRKQAPVKQFQEKAKK 269
Db 908 M--FSARKC--NNFKLDTALSKLPKGRPTLLN---YKNVV----- 941
QY 270 TQSTTTTFFSYTTSAAALNVNTVY-----SITTAARRVSTSTI-----AYRPDSS 315
Db 942 ---TQNYFSPFBGELLATKYKNYLDLNPYQSLKVLGMLQNLATSSALATLVPKTKRAGNK 998
QY 316 FMKSIATQLRLATVWYITLRYQRNPFCEPSNRNVAUSE-FWK--NTHVLIRNE--TPY 370
Db 999 SSKQKIKLNGELUSTELGSTIPQSGKHKTTEKGRMAMFKYLLKTIINSQKIIGNKGWGRF 1058
QY 371 TIYGT-LDMSSLYNETMFVENKTASDNSK-----TTPTSPSGMGPQTFIDPLMDYL 421
Db 1059 NLLITKKDFITLKNYNTLPNTIFSEIQKHQWPIQMTKPIRP-ISYDEVCL----- 1110
QY 422 DLSLFLDEIRNLSRSPPTVNVITPPPH-----RRAVNLSTNS 459
Db 1111 -NYLFSEBITN-QVKLOTLAKLKEIHFNKSYHFNLKNRWLQKLVINASTSKS 1162

RESULT 12
B46629
N:Alternate names: MUC6
C:Species: Homo sapiens (man)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: B46629
R:Toribara, N.W.; Robertson, A.M.; Ho, S.B.; Kuo, W.L.; Gum, E.; Hicks, J.W.; Gum Jr., J.
A:Title: Human gastric mucin. Identification of a unique species by expression cloning.
A:Reference number: A46629; MUID:93194895; PMID:7680650
A:Accession: B46629
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-505 <TOR>
A:Cross-references: UNIPROT:Q14395; GB:L07518; NID:G292045; PIDN:AAB61945.1; PID:G292046
A:Experimental source: Stomach
A>Note: sequence extracted from NCBI backbone (NCBIN:128397, NCBIP:128399)
C:Genetics:
A:Gene: GDB:MUC6
A:Cross-references: GDB:134734; OMIM:158374
A:Map position: 11p15.5-11p15.5
C:Keywords: glycoprotein

Query Match 4.2%; Score 102.5; DB 2; Length 505;
Best Local Similarity 19.4%; Pred. No. 5.1;
Matches 91; Conservative 54; Mismatches 179; Indels 145; Gaps 17;

QY 65 QLETTISTKYNVSKQPVKNLTWNTFFPQYILAGPTQNYISITYLWDFDYSVSTQLRKPAYV 124
Db 61 QAHSSFT--NKTPTSLSHSTSSTHPE-----VAPTSSTTITP---NPTSTRRTPEVA-- 109
QY 125 YSOYNHTAKTITFRPPCGRVPSMTCLSEMLNYSKNDTGEQCGNF--TTFNPM----- 177
Db 110 -----HTNATSRRPP-----PPTTHSPPTGSSPSSGPMATSFKTTITYTPPSLPQ 159
QY 178 ---FFNVPRNWKI-----YVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFYLNV 227

Db 160 TPLTHVPPFFSTSLVTPITHVTPTPQMSTSAVIHSTPTGTIASPTTVKATSTYTAP 219
QY 228 AMSRNLFRVPKVIINGTKLNTMRKLRKQAPVKQFQEKAKTOSTTTPPYFSYTSALN 287
Db 220 LMTATTSISQAHSISIAKTSTSL-----HSHASSTHPEVPTSTVN 263
QY 288 VTNVYVSIITTAARRVSTSTIAYRPDSFMKSIATOLRLD-----ATWVYTLRYRQ 340
Db 264 VTPKSTSRDTSTPVTHTTTSATSSRPPTIHTHSSPTSSPLSSGTPMTATSIKTTTYP- 322
QY 341 NPFCEPSNRT-----AVSEFMKNTHVLI----- 364
Db 323 ---TPSHQPQTLTHVPPFFSTSSVTSTHTVITPTHAQMSSTASIHSTPTGTVPPLTR 378
QY 365 -----KNETPYTYIGTLDMSLYNETMFVENKTASDNS-----KTTPTSPSGMGPQ 411
Db 379 MPTGSTRGTGPMW--GHIIQTSKAHNSFSAKTSTLSHSHASSTHPEPTTSTT----- 431
QY 412 TFDPLMDYLDLFLDEIRNLSRSPYYNLTTPPEHRAVNLSTNSL 460
Db 432 -----NITPKSTSACTSTPVAH---TTLATSSRL 457

RESULT 13
T38863
hypothetical protein SPAC4G9.04c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T38863
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21732
A:Accession: T38863
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-638 <CON>
A:Cross-references: UNIPROT:Q10237; EMBL:Z69727; PIDN:CAA93554.1; GSPDB:GN00066; SPDB:SP.
A:Experimental source: strain 972h-; cosmid c4G9
C:Genetics:
A:Gene: SPDB:SPAC4G9.04c
A:Map position: 1
A:Introns: 50/3

Query Match 4.2%; Score 102.5; DB 2; Length 638;
Best Local Similarity 19.0%; Pred. No. 7;
Matches 83; Conservative 73; Mismatches 160; Indels 121; Gaps 17;

QY 55 LDQLKLEILRQLQLETTISTKYNVSKQPVKNLTWNTFFPQYILAGPTQNYISITYLWDFDYS 114
Db 1 MDLVELDVLSALE-----DLTFNSK-----PI-IHTITYI----- 29
QY 115 TQLRKPAYVYSQVNHAKTITFRPPCGRVPSMTCLSEMLNYSKNDTGEQCGNFPTTF 174
Db 30 AOENEP--YAISIVNAIEKHIQKCPNCC-KLPALYLLDSISKNLGAPYTYFFGLHLFTF 86
QY 175 NPMFNV--PRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSFVLVNMARNL 233
Db 87 MSAYTVVEPRLEK-----LDLIATWKQR-----PPNSSSL 118
QY 234 FRVPKVIINGTKLNTMRKLRK-----RKQAPVKQFQEKAKTOSTTTPPYFSYTSALNVT 289
Db 119 EPVFSPIVTAIKIENALLKYKSTILRHQSPL-----LANTSISFSAIDANIN 166
QY 290 TNVYVSIITTAARRVSTSTIAY-----RPDSSFMKSIATOLRLATVWYTLRYRQ 340
Db 167 SYSEFSDPASYSKESLSPVPGFOHISGTSPSGFI--TLDSLLSDVNRNIVTEQARFIK 224
QY 341 NPFCEPSNRTAVSEFMKNTHVIRNETPYTYIGTLDMSLYNETMFVENKTASDNSKT 400
Db 225 NPYDNMAKRPFIILLQKN-----VLSSSALPYDOLLAIKNQLAQLEKPA 269
QY 401 TPTSPSGMGPQTFIDPLMDYLDLFLDE-----IRNFSLSRSP---TYVNL 444

Search completed: December 7, 2004, 15:21:27
Job time : 47 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:03:02 ; Search time 154 Seconds
(without alignments)
1083.176 Million cell updates/sec

Title: US-09-942-146A-1

Perfect score: 2451

Sequence: 1 MGRKEMVRDVPKMFVLISI.....PEHRRVNLSTNSLSLMMWLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04.*

- 1: geneseqp1980a.*
- 2: geneseqp1990a.*
- 3: geneseqp2000a.*
- 4: geneseqp2001a.*
- 5: geneseqp2002a.*
- 6: geneseqp2003a.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004a.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2451	100.0	465	8	AD015374
2	2419	98.7	465	7	AD015374
3	111.5	4.5	2590	8	AD028932
4	110.5	4.5	1304	6	ABP72189
5	110	4.5	2172	4	ABP67958
6	109	4.4	1091	8	ADN05348
7	108.5	4.4	1224	3	ABP18258
8	108	4.4	1327	5	ABP97286
9	107.5	4.3	1168	2	AAW16326
10	106.5	4.3	1167	2	AAW10653
11	106.5	4.3	1167	2	AAW31504
12	105	4.3	452	7	ADP07589
13	103.5	4.2	1062	2	AAW22721
14	101.5	4.1	1042	6	ABP57376
15	101	4.1	2481	8	AD028930
16	101	4.1	2570	4	ABG06375
17	100.5	4.1	1224	6	ABP57375
18	100.5	4.1	4315	5	ABP43908
19	100.5	4.1	5703	8	ADL23265
20	99.5	4.1	501	2	AAW82571
21	99.5	4.1	978	2	AAW17509
22	99.5	4.1	1349	5	ABP47323
23	99	4.0	903	4	ABP64101
24	98.5	4.0	622	6	ABP77408
25	98.5	4.0	622	6	ABU37421

26	98.5	4.0	4134	2	AAV31946
27	98	4.0	234	3	AAG06725
28	98	4.0	234	3	AAG40390
29	98	4.0	313	3	AAG06724
30	98	4.0	313	3	AAG40389
31	98	4.0	393	3	AAG40388
32	98	4.0	393	3	AAG06723
33	98	4.0	393	5	ABP91164
34	97.5	4.0	638	7	ADL90059
35	97.5	4.0	992	8	ADP04643
36	97.5	4.0	1365	6	ABP72194
37	97.5	4.0	2404	4	ABP69209
38	97	4.0	621	6	ABU37812
39	97	4.0	1007	4	ABP65680
40	97	4.0	2000	4	AAAB61853
41	96.5	3.9	347	7	ADJ68384
42	96.5	3.9	698	8	ADP99077
43	96	3.9	706	2	AAZ75647
44	96	3.9	859	5	ABP27958
45	96	3.9	1296	5	ABG66756

ALIGNMENTS

RESULT 1

ID	AD015374	standard; protein; 465 AA.
XX	AD015374;	
AC	AD015374;	
XX		
DT	01-JUL-2004	(first entry)
XX		
DE	CMV glycoprotein O (gO) polypeptide.	
XX		
KW	CMV; glycoprotein O; gO; glycoprotein H ; glycoprotein L; glycoprotein B;	
KW	CMV infection; tumour; cancer; virucide.	
XX		
OS	Cytomegalovirus.	
XX		
PN	US2004013682-A1.	
XX		
PD	22-JAN-2004.	
XX		
PF	26-JUN-2002; 2002US-00942146.	
XX		
PR	29-JUL-1999; 99US-0146180P.	
PR	28-JUL-2000; 2000US-00627986.	
XX		
PA	(COMP/) COMPTON T.	
PA	(HUBE/) HUBER M T.	
XX		
PI	Compton T, Huber MT;	
XX		
DR	WPI; 2004-121532/12.	
XX		
PT	Designing new anti-CMV drug by analyzing binding of glycoprotein O to	
PT	glycoprotein O receptor, designing candidate drug that would	
PT	competitively interfere with glycoprotein O binding to glycoprotein O	
PT	receptor.	
XX		
PS	Disclosure; Fig 3; 7pp; English.	
XX		
CC	The invention relates to a method of designing a new anti-cytomegalovirus	
CC	(CMV) drug, involving analysing the binding of glycoprotein O to a	
CC	glycoprotein O receptor, designing a candidate drug that would	
CC	competitively interfere with glycoprotein O binding to the glycoprotein O	
CC	receptor and showing that the candidate drug competitively inhibits	
CC	glycoprotein O binding to the glycoprotein O receptor. The invention also	
CC	relates to a method of screening involving determining whether a	
CC	candidate drug interferes with a glycoprotein O-containing complex	
CC	binding to a cell surface. The evaluation involves analysing the	
CC	inhibition of major early protein of the human cytomegalovirus (HCMV).	

CC The invention also relates to a vaccine useful for diminishing CMV
 CC infection and comprising at least a fragment of the glycoprotein O
 CC polypeptide in combination with a carrier. The vaccine comprises at least
 CC a fragment of a glycoprotein chosen from CMV glycoprotein H, CMV
 CC glycoprotein L or CMV glycoprotein B, and is useful for vaccinating a
 CC patient against CMV infection. The drug and the vaccine are useful for
 CC treating CMV infections in tumours. This sequence represents the CMV
 CC glycoprotein O (gO) polypeptide of the invention.
 XX
 SQ Sequence 465 AA;

Query Match 100.0%; Score 2451; DB 8; Length 465;
 Best Local Similarity 100.0%; Pred. No. 3.3e-218;
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60
 DB 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60
 QY 61 EILRQLETTISTKYNVSKQPVKNLTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
 DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
 QY 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFFN 180
 DB 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFFN 180
 QY 181 VPRWNTKLVGPTKVNVDSTQTYIFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKYI 240
 DB 181 VPRWNTKLVGPTKVNVDSTQTYIFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKYI 240
 QY 241 NGTKLKNTRKLRKQAPVKEQFEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300
 DB 241 NGTKLKNTRKLRKQAPVKEQFEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300
 QY 301 RVSTSTIAYRPDSSFMKSIMATQLRDLATWYVYTLRYQRNPFCPEPSRNRRTAVSEPMQNT 360
 DB 301 RVSTSTIAYRPDSSFMKSIMATQLRDLATWYVYTLRYQRNPFCPEPSRNRRTAVSEPMQNT 360
 QY 361 HVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPMGFORFTIDPLWDY 420
 DB 361 HVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPMGFORFTIDPLWDY 420
 QY 421 LDSLFLDLEIRNFSRLSPYVNLTPPEHRAVNLSTSNLWMLQ 465
 DB 421 LDSLFLDLEIRNFSRLSPYVNLTPPEHRAVNLSTSNLWMLQ 465

RESULT 2
 ADB81330
 ID ADB81330 standard; protein; 465 AA.
 XX
 AC ADB81330;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE Human cytomegalovirus glycoprotein O polypeptide.
 XX
 KW glycoprotein O; gO; anti-cytomegalovirus; CMV; AIDS;
 XX bone marrow transplant; glycoprotein H; glycoprotein L; gene therapy.
 XX
 OS Human herpesvirus 5.
 XX
 FN US6569616-B1.
 XX
 PD 27-MAY-2003.
 XX
 XX 28-JUL-2000; 2000US-00627986.
 XX
 PR 29-JUL-1999; 99US-0146180P.
 XX
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 XX

PI Compton T, Huber MT;
 XX WPI; 2003-605462/57.
 DR
 XX Screening candidate anti-cytomegalovirus (CMV) drugs for the ability to
 PT block CMV entry into a host cell comprises determining whether a
 PT candidate drug interferes with glycoprotein O binding to glycoprotein O
 PT receptor.
 XX
 XX Disclosure; Fig 3; 8pp; English.
 XX
 CC This invention relates to a novel method of designing a new anti-
 CC cytomegalovirus (CMV) drug that has the ability to block CMV entry into a
 CC host cell. CMV is a member of the Herpesviridae family and is an
 CC opportunistic pathogen responsible for serious clinical disorders in
 CC immunosuppressed patients, such as persons with AIDS or those with recent
 CC organ or bone marrow transplants. The present invention describes a new
 CC drug target that comprises a CMV viral component designated glycoprotein
 CC O (gO), whereby the candidate drug works by interfering with gO binding
 CC to its receptor, where glycoprotein L. Accordingly, using gene therapy or
 CC glycoprotein H and glycoprotein L. Furthermore, it is possible to block
 CC developing antibodies to gO that act as vaccines. virus infection by
 CC method is useful in identifying alternative drug targets and immunogens
 CC that elicit protective activity against CMV infection. This polypeptide
 CC is the human cytomegalovirus glycoprotein O amino acid sequence of the
 CC invention.
 XX
 SQ Sequence 465 AA;

Query Match 98.7%; Score 2419; DB 7; Length 465;
 Best Local Similarity 99.4%; Pred. No. 3e-215;
 Matches 463; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60
 DB 1 MGRKEMVRDVPKMFVLISISFLVSVFVCKVMSKALYNRPWGLVLSKIGYKLDQLKL 60
 QY 61 EILRQLETTISTKYNVSKQPVKNLTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
 DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTPEPQYVILAGPIQNYISITVLMWDFYSTQLRKP 120
 QY 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
 DB 121 AKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
 QY 180 NVRWNTKLVGPTKVNVDSTQTYIFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 239
 DB 181 NVRWNTKLVGPTKVNVDSTQTYIFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 240
 QY 240 INGTLKNTRKLRKQAPVKEQFEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 299
 DB 241 INGTLKNTRKLRKQAPVKEQFEKAKKTOSTTTPYFSYTTSAALNVTNTVYSITTA 300
 QY 300 ARVYSTIAYRPDSSFMKSIMATQLRDLATWYVYTLRYQRNPFCPEPSRNRRTAVSEPMKN 359
 DB 301 ARVYSTIAYRPDSSFMKSIMATQLRDLATWYVYTLRYQRNPFCPEPSRNRRTAVSEPMKN 359
 QY 360 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPMGFORFTIDPLWD 419
 DB 360 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPMGFORFTIDPLWD 419
 QY 420 YLDSLLFLDLEIRNFSRLSPYVNLTPPEHRAVNLSTSNLWMLQ 465
 DB 420 YLDSLLFLDLEIRNFSRLSPYVNLTPPEHRAVNLSTSNLWMLQ 465

RESULT 3
 ADO28932
 ID ADO28932 standard; protein; 2590 AA.
 XX
 AC ADO28932;
 XX

DT 29-JUL-2004 (first entry)
 XX Mouse novel GPCR PGR17, SEQ ID NO:31.
 XX
 XX G protein-coupled receptor; GPCR; drug screening; diagnosis;
 KW transgenic mouse; neurological disorder; adrenal gland disorder;
 KW colon disorder; intestinal disorder; cardiovascular disorder;
 KW muscular disorder; blood disorder; immune disorder; bone disorder;
 KW joint disorder; metabolic disorder; nutritive disorder; cancer;
 KW kidney disorder; liver disorder; lung disorder; breast disorder;
 KW ovary disorder; uterus disorder; prostate disorder; testis disorder;
 KW skin disorder; stomach disorder; pancreas disorder; spleen disorder;
 KW thymus disorder; thyroid disorder; antiparkinsonian; animanic;
 KW cycostatic; antinflammatory; vasotropic; antidiarrhoeic; antidiabetic;
 KW CNS; central nervous system; respiratory; antidiarrhoeic; antidiabetic;
 KW virucide; hepatotropic; antibacterial; antianemic; antiseborrheic;
 KW dermatological; antitumor; antithyroid; antiallergic; anorectic;
 KW immunosuppressive; nephrotropic; gene therapy; GPCR modulator; mouse;
 KW murine; receptor.
 XX
 OS Mus musculus.
 XX
 XX WO2004040000-A2.
 XX
 XX 13-MAY-2004.
 XX
 XX 09-SEP-2003; 2003WO-US28226.
 XX
 XX 09-SEP-2002; 2002US-0409303P.
 XX
 XX 09-APR-2003; 2003US-0461329P.
 XX
 XX (PRIM-) PRIMAL INC.
 XX
 XX Gaitanaris GA, Bergmann JB, Gragerov A, Hohmann J, Li F;
 PI Madisen L, McIlwain KL, Pavlova MN, Vassilatis D, Zeng H;
 PI
 XX WPI; 2004-390329/36.
 DR
 DR N-PSDB; ADO28933.
 XX
 XX Novel mammalian G protein coupled receptors, useful for identifying
 PT compounds that modulates diagnosing and treating disease condition
 PT associated with GPCR dysfunction e.g. autoimmune diseases, angina
 PT pectoris, Parkinson's disease.
 XX
 XX Claim 1; SEQ ID NO 31; 542pp; English.
 XX
 XX The invention relates to human and mouse G protein-coupled receptors
 CC (GPCRs) and nucleic acids encoding them. The invention also relates to
 CC sequences at least 90% identical to the GPCR proteins and nucleic acids
 CC of the invention; methods of treating, preventing or diagnosing diseases
 CC associated with GPCRs of the invention; methods of screening for
 CC compounds useful in the treatment of GPCR-related diseases; a transgenic
 CC mouse comprising a GPCR gene of the invention; a mouse comprising a
 CC mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived
 CC from the transgenic mice; kits comprising several mice, each of which has
 CC a mutation in a different GPCR gene of the invention; and kits comprising
 CC probes which hybridise to GPCR polynucleotides of the invention. The
 CC invention further discloses variants of the GPCR polypeptides and vectors
 CC comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may
 CC be used in the diagnosis, treatment or prevention of a wide variety of
 CC diseases including neurological disorders (e.g., Alzheimer's disease,
 CC depression, diabetic neuropathy, Parkinson's disease or schizophrenia);
 CC disorders of the adrenal gland; disorders of the colon or intestine
 CC (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel
 CC syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or
 CC myocardial infarction); muscular disorders; blood disorders (e.g.,
 CC anaemia or leukaemia); immune disorders (e.g., autoimmune disorders or
 CC AIDS); bone and joint disorders (e.g., osteoarthritis, rheumatoid
 CC arthritis, gout or osteoporosis); metabolic or nutritive disorders (e.g.,
 CC obesity, enzyme deficiency-related diseases or vitamin deficiency-related
 CC diseases); and disorders of the kidney, liver, lung, breast, ovary,
 CC uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and
 CC thyroid (e.g., cancers). The present sequence represents a GPCR of the

CC invention. Note: The full sequence data for this patent did not form part
 CC of the printed specification; those sequences not shown were obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 2590 AA;
 XX
 XX Query Match 4.5%; Score 111.5; DB 8; Length 2590;
 XX Best Local Similarity 23.3%; Pred. No. 3.4;
 XX Matches 60; Conservative 39; Mismatches 116; Indels 43; Gaps 11;
 XX
 QY 225 LVNAMSRLFRVP-KYINGTKLNTMKLRKQAPVKEQPEKKAKTKQS--TTTPYRSYT 281
 DB 1725 LYNANSR--ATIPANEUSTPSSDNFYTLNQDSPTLTNSKVTPRPTESVKSTPTHLSFD 1782
 QY 282 TSAALNVTNVTYS---ITTAARR---VSTSTI-----AYRPDSFPMKSIMATQLRDLA 329
 DB 1783 T-RKMNLTETKSGPCVTTTPVLYPLWTQTSTAPPLTSHLYSPHSTRAKKPLASQMAEYP 1841
 QY 330 TWVYTLRYQNPFCEPSRNRATAVSFPMKNTVLIIRNETPYTYIYGTLDMSLYYNETMFV 389
 DB 1842 AWATGITPSITQALLTTSRNTQV-----EDSPFPVPTTKDQGLSTSTFLE 1889
 QY 390 ENKTASDSNKTTPSPSMGQRTFIDPLWDYLDLSLLFLD-EIRNFSLSRPT-----YVN 442
 DB 1890 SSLRTTGADSSDMSRMSFGRTSISP-----SLTRHDLISGLTSSPTNTSPWSKVP 1943
 QY 443 LTPPEHRRAVNLSTNSL 460
 DB 1944 VTSESHLTFPSKTLDSV 1961
 XX
 XX RESULT 4
 XX ABP72189
 ID ABP72189 standard; protein; 1304 AA.
 XX
 XX AC ABP72189;
 XX
 DT 22-APR-2003 (first entry)
 DE Plasmodium berghei modular secreted protein PbSR.
 XX
 XX PbSR; protein secretion; malaria; diagnosis; vaccine; antimalarial;
 KW protozoacide.
 XX
 XX Plasmodium berghei.
 XX
 XX Key Location/Qualifiers
 FT Peptide 1..22
 FT Protein /label= Signal_peptide
 FT /label= Mature_protein
 XX
 XX WO2003004524-A2.
 XX
 XX 16-JAN-2003.
 XX
 XX 02-JUL-2002; 2002WO-GB003045.
 XX
 XX 02-JUL-2001; 2001GB-00016185.
 XX
 XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
 XX
 XX Claudianos C, Crompton TK, Dessens JT, Sinden RE, Trueman HE;
 XX
 XX WPI; 2003-210339/20.
 DR N-PSDB; ABZ58196.
 XX
 XX New secreted proteins from malarial parasites (e.g. Plasmodium
 PT falciparum), useful for detecting or diagnosing malaria, or as a vaccine
 PT or medicament for the treatment and/or prophylaxis of malaria.
 XX
 XX Claim 1; Fig 1; 67pp; English.
 PS

XX CC The present sequence is the protein sequence of novel secreted protein
CC PHSR from the mouse malaria parasite, Plasmodium berghei. The invention
CC provides novel modular secreted proteins from Plasmodium falciparum
CC (PfSR), P. berghei (PbSR) and Plasmodium yoelii (PySR) that are post-
CC translationally modified in sporozoites. These include Limulus clotting
CC factor, lipid binding, scavenger receptor cysteine-rich and pentraxin
CC domains. Homologues were identified in Plasmodium knowlesi (PKSR),
CC Plasmodium chabaudi (PCSR), Cryptosporidium parvum (CpSR), Toxoplasma
CC gondii (TgSR) and Theileria parva (TpSR). PbSR, PfSR and PySR are from a
CC unique and previously unrecognized family containing domains associated
CC with binding and modulating host proteins involved in immunity. They are
CC the first proteins from Plasmodium which are secreted and which show a
CC gene disruption phenotype in both vertebrate and mosquito, making them
CC highly useful for the production of malaria vaccines. Transgenic PbSR
CC null mutants have attenuated growth in immunocompetent mice. They fail to
CC produce sporozoites in Anopheles stephensi mosquitoes and consequently
CC are dead-end parasites in the midgut of the mosquito. The novel
CC Plasmodium proteins, and the nucleic acids encoding them, are used in
CC claimed methods for the detection/diagnosis of malaria, and are also used
CC in vaccine compositions. Agents capable of antagonising, inhibiting or
CC interfering with the function or expression of the secreted protein are
CC used in the treatment and/or prophylaxis of malaria

XX SQ Sequence 1304 AA;
Query Match 4.5%; Score 110.5; DB 6; Length 1304;
Best Local Similarity 20.2%; Pred. No. 1.6;
Matches 79; Conservative 42; Mismatches 129; Indels 141; Gaps 15;
QY 122 KYVSYQYNNHTATITFRPPCGRPVSMCLSEML-----NVSKND-TGEGCGGNF--- 171
DB 873 KYIIRLSNFTF-IIHFIPNSGKKNKWTILSHSLCEGISIDBENELVIBQNCNPHLVK 931
QY 172 TTENPMF-----ENVPRWNTKLYVGPVKVNDVDSQ-----TIY 203
DB 932 TIFPKFEPCHLVLYINKPKNSISLYNQKINLEKMKFDTFLNGDLITGRSNKQATDY 991
QY 204 FLG-----ITALLRYAQRNCTHISFVLVNMGRNLFRVPKYINGTKLKNYWRKL 252
DB 992 FIGDINFVKYIKYTEQIKESYDLSVLSNLYNDGMSGN-----RDINTKKTQNKTKN 1046
QY 253 KKKQ-----APVKEQEKAKKTQSTTTPVFSYTTSAALNVTTN-----VTYS 295
DB 1047 NRKTDIGRCDITPCSKNTVVK-KNVQINTBEFLDCSDNLLSERFNGKIGQFLVSCLED 1105
QY 296 ITTAARRVSTSIAYRPDSSFMKSIATQLRDLATWVYTTLYRONPFCPSRNRATVSE 355
DB 1106 CTNSKIVYKGSNNYTFDTSICKAVM-----HSGIMH 1137
QY 356 FMKNTHVLIRNETPYTIYGTLDMSLLYNNETMFVENKTAASDNKTTPTSPSMGFQRTFD 415
DB 1138 KTRNTH-----KNEHEDNNKNTSNS-----FLIKIVE 1166
QY 416 PLWDYLDLSLFL-----DEIRNFSLS 437
DB 1167 GLTEYKSRGHGIVSKPEKQSLRSFSLFS 1197

RESULT 5
ABB67958
ID ABB67958 standard; protein; 2172 AA.
XX
AC ABB67958;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 30666.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.

XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX DR N-PSDB; ABL12061.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX interactions.
XX PS Disclosure; SEQ ID NO 30666; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at fcp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2172 AA;
Query Match 4.5%; Score 110; DB 4; Length 2172;
Best Local Similarity 19.4%; Pred. No. 3.6;
Matches 87; Conservative 62; Mismatches 174; Indels 126; Gaps 17;
QY 67 ETTISTKYNVSK-----QPVKNLTMTNTEFFQYIYLAGPIQNIYSITYLWFDYFTQ-----LR 118
DB 1450 KTSSTSYTTTSTTYIQPSTSYLNAEVTSH---PTTSTYEVNTIITDESKIKDMFNLI 1506
QY 119 KPARYVSYQYNNHTAK---TITFRPPCGRPVSMCLSEMLNVSKRNDTGEQCGNETTFN 175
DB 1507 KGIEDLIQYNGKHKEFTTSLTGKNSRPTWTYSSAGGITSH-----SFTTIS 1557
QY 176 PMFNPVPRWNTKLYVGP-----KVNVDSDQIYFLGLTALLRYAQRNCTHSHFYLVA 228
DB 1558 PKTVTSSSHSTTESPSHPTTNTSYEVTVPDQD-----SHKFKVMKN 1599
QY 229 MSRNL-----FRVPKYINGTKLNTMRKLKRQAPVKQEFKKAKK----- 269
DB 1600 LIKELKGLIEHFSDTKNTITSTYTGDSKGFQPTATTPDSTDRSSSQIPFPSTTTTN 1659
QY 270 TQSTTTPYFS-YTTSAAALNTVNTVYISITTAARVSTSIAYR-----PDSSFMKSIAT 323
DB 1660 LQSTITNHLSTESTSQNLTTLSIT---GTTPNLTSTHDFKQIQNDSDSKSLKHU--- 1713
QY 324 QLRLATWVVT-----TLRYRONPFCPSRNRATVSEFMKNTHVLIRNETPY 370
DB 1714 -IKILSDLVSLNDKNKIKSTSLTSTFNSSINITNSTVASTSTSTTESMAIHSTPY 1772
QY 371 TIYGTLDMSLLYNNETMFVENKTAASDNKTTPTSPSMGFQRTFDPLWDYLDLSLLFDEI 430
DB 1773 --LSTREITVFQND---VENINPIISNNSTVTT----- 1802
QY 431 RNFSLRSPYTNLTTPPEHRRAVNLSTNS 459
DB 1803 ---NIHSSYTTTSTTDNTTSYPNLSTTES 1828

RESULT 6

ADN05348 standard; protein; 1091 AA.
 AC ADN05348;
 XX
 DT 01-JUL-2004 (first entry)
 DE Antipsoriatic protein sequence #846.
 XX antipsoriatic; gene therapy; psoriasis; diagnosis.
 KW Homo sapiens.
 OS
 XX WO2004028479-A2.
 XX
 XX 08-APR-2004.
 XX
 XX 25-SEP-2003; 2003WO-US030907.
 XX
 XX 25-SEP-2002; 2002US-0414006P.
 XX
 XX (GETH) GENENTECH INC.
 XX Bodary S, Clark H, Jackman J, Schoenfeld J, Williams PM, Wood WI;
 PI Wu TD;
 PI WPI; 2004-305105/28.
 DR N-PSDB; ADN05347.
 XX
 XX New PRO nucleic acid or polypeptide, useful for preparing a
 PT pharmaceutical composition for diagnosing or treating psoriasis in a
 PT mammal.
 XX
 XX Claim 9; SEQ ID NO 1742; 3069pp; English.
 XX
 XX The invention relates to novel polynucleotide and polypeptides for
 CC treating psoriasis or a sequence having at least 80% identity to the
 CC above sequences. The nucleic acid is useful for preparing a composition
 CC for diagnosing or treating psoriasis in a mammal. This sequence
 CC corresponds to one of the polypeptides of the invention.
 XX
 XX Sequence 1091 AA;
 Query Match 4.4%; Score 109; DB 8; Length 1091;
 Best Local Similarity 20.7%; Pred. No. 1.7;
 Matches 84; Conservative 60; Mismatches 167; Indels 94; Gaps 20;
 Qy 55 LDQLKLEILRQLETTISTKYNVSKQPVKNL-TMNTFPPQYVILAGPIONYSITYLWDFD- 112
 Db 540 LDPLDAELNDIKVEIRNKMIDGESGKTFRTLVKSQDERVIDKG-----NRTVTWTFVN 594
 Qy 113 ---YSTQLRKPAKYVSYNHTAK---TITFRPPCGRPVMTCLSEMLNVSKNDTGEQ 166
 Db 595 GTDYSALVLP---TYSFYIKAKLEETITQARSKGRMKD-----SETLKPDNFEESG-- 645
 Qy 167 GCGNFTTPNPFNVPRNNTKLYGPKVNDVSDTIYFLGLTALLRVAQRNCTHSFVLV 226
 Db 646 ----YTFIAP-----RDYCNLDKIS-DNNTFELNFPNEDRTPNPNPCNADLI 690
 Qy 227 -----NAGRNLPFRPKYINGTKLKNVM-----RKLKRKQAPVKEQ 262
 Db 691 NRVLIDAGFTNELQVNVSKQNIKGKVPVTDWMDYQSLSQGRWKLARKPETVEDS 750
 Qy 263 PEKAKKTQS---TTPPYSTYSAL--NVTNVTYSITTAARVSTIAVRPD-SFPM 317
 Db 751 FYKSLDNDNVTFAPYFNKSGPGAYESGIMVSKAVEIYIQGLKLPVAVGKIDVNSWI 810
 Qy 318 KSINATOLDLATWYTTTLRYQRNPFCEPSNRRTAVSEFM-----KNTHVLIRNETPTT 371
 Db 811 ENFTKTSIRDPCA-----GPVCKRN-----SDVMDCVLDDGGFLMANHDDYT 856
 Qy 372 ----YGTLDWSSLY--NETMFVENKTASDSNKTTP-TSPSMG 408

Db 957 NOIGRFFGEIDPSLRHLVNISVYAFNKSXDYQSVCEPGAAPKQG 901

RESULT 7

AAB18258 standard; protein; 1224 AA.
 XX AAB18258;
 XX
 DT 07-NOV-2000 (first entry)
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:115.
 XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX Plasmodium falciparum.
 OS
 XX WO200025728-A2.
 XX
 XX 11-MAY-2000.
 XX
 XX 05-NOV-1999; 99WO-US026796.
 XX
 XX 05-NOV-1998; 98US-0107131P.
 XX (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX Hoffman S, Carucci D, Gardner M, Venter JC;
 PI WPI; 2000-365347/31.
 XX
 XX Proteins encoded by chromosome 2 of the human malarial parasite, and in the
 PT diagnosis of P.falciparum infection.
 XX
 XX Disclosure; Page 277-280; 577pp; English.
 XX
 XX The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II). (I) and
 CC (II) are useful for the development of vaccines against P. falciparum
 CC infection. (I) and polyclonal antisera or a monoclonal antibody raised to
 CC immunogens comprising the sequences of (I), are useful in the detection
 CC of infection with P. falciparum. Furthermore, (I) (especially when they
 CC are fusions or secreted or membrane proteins) can aid the identification
 CC of drugs to treat or prevent P. falciparum infection, or they can be used
 CC to identify drug resistance in P. falciparum. Sequencing of the
 CC Plasmodium chromosome 2 and the subsequent identification of proteins
 CC encoded by it will help to expand our understanding of parasite biology,
 CC a process hampered by the complexity of the parasitic lifecycle, and
 CC provide new targets for vaccine and drug development. Parasite resistance
 CC to drugs and mosquito resistance to insecticides have led to a resurgence
 CC of malaria in many parts of the world, and there is a pressing need for
 CC vaccines and new drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352
 CC represent nucleotide and protein sequences given in the present
 CC invention, but which are not specifically mentioned within the
 CC specification
 XX Sequence 1224 AA;
 Query Match 4.4%; Score 108.5; DB 3; Length 1224;
 Best Local Similarity 21.7%; Pred. No. 2.2;
 Matches 78; Conservative 47; Mismatches 116; Indels 119; Gaps 19;
 Qy 73 KYNVSKQPVKNLTWNTF--FPQYVI-LAGPIONYSITYLW-----PDF----- 112
 Db 82 KYRKNNNNVKNMILKDELDILYDYNHLSNHLNHDINFIYSSNNIFNLCKNNKPKYFNS 141

113 -YSTQLRPAK-----YYSQYNTAKTITPRPPCGRVSMTCLEMLN-----VSKRN 161
142 KNSNEIKDQKQKVVYNNIHHYTKGKNFYSNPT-EVNNLSLNNKNSLYISFRK 200
162 DTGQGGCGNFTFNPMPFNVRWNTKLYVGTQKVVDSQTIYFGLTALLRLVAQRNCTH 221
201 DT-----SNFN-----PSCDKNT-----TFSPKNCILH 223
222 SFYLVNAMS-----NLFRVPKYNGTKLKNMRLKR-KQAPVKEQFEKAKKTQSTT 275
224 E---SNPSSTCYPNVNTIPLAIN--LLANNVDDISPIHPLPSESSTSASTSAST 278
276 PYFSYTTSAALNVTNTYTSITTAARRVSTSTIAVRDSSPMKSMATQLDLATVYTT 335
279 ---SASTSASTSVSTSVSTSA---SITMNSPRPSDNHISNFPPLSRESRATEQVNR 332
336 LRY-----RONPCEPSRNRNAVSEFMKNTHVLRNETPYTI 372
333 LYFPVNDVTSKDPNPNNELTSNNPKHEPICEETRN-----DNHG--IRNSIYPL 382

RESULT 8
ABB97286
ID ABB97286 standard; protein; 1327 AA.
XX
AC ABB97286;
XX
DT 28-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 554.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antiinfectivity; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
FN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32472.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 554; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention
XX Sequence 1327 AA;
XX

Query Match 4.4%; Score 108; DB 5; Length 1327;
Best Local Similarity 22.0%; Pred. No. 2.8;
Matches 108; Conservative 59; Mismatches 221; Indels 102; Gaps 23;
QY 20 ISPLVSNFKVMSKALYNRPWRGLVLSKIGYKLDQLKLEILRQLETTISTKY---NV 76
DB 266 IQFLMT---NEETVDKA---PPHSKIGLEKRRKRMDVSK--ITRYTEOCFSDSNCPNK 317
QY 77 SK-QPVKNLTMTNTEFPQYVILAGPIQNYSTIYLWDFYSTQLRKPQAKYVYSQVNHAKTI 135
DB 318 SKMQEVDVLEQNEE-----LQAVDSQKIALSKVPESTDEDLSDVAFQHLIYN----- 366
QY 136 TFRPPPCGR-----VPSMTCLEMLNVSXKNDTGEQCGNFTFNPMPFNVRWNTKLYVG 191
DB 367 ---PDKGEBESSPVHTSTFLSNTLK-KKCEESDESPPATFSTEEPSPYPCCTKCNVNF--- 419
QY 192 PTKVNVDSQTIYFGLTALLR-----YAQRNCTHSFYLVNAMSRLFRVPKYINGTKL 245
DB 420 REKKHLHRHMMYHLGDNHSHFRHLNVRPYACRECGRTFRDRNSLLKHM--IIHOERRQKL 477
QY 246 KNTMRKLK-----RQAP-----VKEQFEKAKKTQSTTTPY-----FSYTT 283
DB 478 MEIRELKELODEGRSARLQPCQVGTNCPKTFVQHAKEKDKRYCCCEECNFMVATE 537
QY 284 AALNVTNTYTSITTAARRVSTSTIAYR--PDSSFMKSMATQLDLATVYTTLYRON 341
DB 538 NELECHRGIAHGAVVKCPMV-TSDIAQRKTQKTFMKDSVVGSKKSATYI-----CKMC 591
QY 342 PCEPSRNRNAVSEFMKNTHVLRNETPYTIYGTLDMSLYYNETMFVENKTASDSNKT 401
DB 592 PF-----TTSKSVLKXHTEYLHSSSCVDSFGSPLGLDK---RNDILEEPVSDSTKTL 643
QY 402 PTSFGMGFORFTDPLWDYLDLSLLFLDEIRNF--SLRSPTYVNLTPPEHR----- 449
DB 644 TKQOSTTTPK-----NSALKQDVKTFTGSTSSQSSFSKIHKPHRIQARKASIAQ 693
QY 450 RAVNLTSTNS 459
DB 694 SGVNMCMQNS 703

RESULT 9
AAW16326
ID AAW16326 standard; protein; 1168 AA.
XX
AC AAW16326;
XX
DT 17-AUG-1997 (first entry)
XX
DE Nematocidal toxin 167P.
XX
KW Nematocide; pesticide; biological control; corn rootworm; Diabrotica;
KW crystal protein; CryV; endotoxin.
XX
OS Bacillus thuringiensis.
XX
FH Key Location/Qualifiers
FT Misc-difference 1168
FT /note= "the methionine residue at C-terminal position
FT 1168 is additional to the sequence deduced from the 167P
FT gene"
XX
FN WO9712980-A1.
XX
PD 10-APR-1997.
XX
PF 01-OCT-1996; 96WO-US015730.
XX
PR 06-OCT-1995; 95US-00540104.
XX 21-MAR-1996; 96US-00620717.
XX (MYCO) MYCOGEN CORP.
XX

```

PI Feitelson JS;
DR WPI; 1997-226223/20.
DR N-PSDB; AAT62498.
XX
XX New Bacillus thuringiensis isolates - polynucleotide sequences encoding
XX toxins useful for controlling nematode and coleopteran pests.
XX
XX Disclosure; Page 31-35; 46pp; English.
XX
XX The Bacillus thuringiensis 167P pesticidal toxin (AAW16326) can be
XX expressed in recombinant microorganisms, e.g. B.t., E. coli or
XX Pseudomonas, for use in the biological control of pests, or expressed in
XX transgenic plants to improve pest resistance. The amino acid sequence of
XX 167P was deduced from the 167P gene coding sequence (AAT62498)
XX
XX SQ Sequence 1168 AA;

Query Match 4.4%; Score 107.5; DB 2; Length 1168;
Best Local Similarity 20.9%; Pred. No. 2.6;
Matches 97; Conservative 68; Mismatches 131; Indels 169; Gaps 26;

Qy 90 PPOYY-ILAGPIQNYSTIYLFDFYSTOLRKPAPKYVSYQVNHAKTITFRPPPCGRVPSM 148
Db 8 YPSHNVLHPIRLDS-----FFDPP-VETFKDLKGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTFNPMFNVPRWNTKLVGPTKVND-----SQ 200
Db 49 DPLKQHLQIAW--DTSQNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYFGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYNG- 242
Db 89 VVPFINMFVDFIPKLFGRGSGQNAQAQFFELIIEKVKELVDEDFRNFNLNLLYLDGM 148
Qy 243 -----TKLKNTR-----KLKRKQAP-----VKQPEKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQGLMDQTPTACTPTDHLISVRESF-KDARTTIETALPH 207
Qy 278 FS-----YTTSALNVNTNVTYSITTAARRVSTSIAYRPD 313
Db 208 FKNPMLSTNDNTPDPSNTVLLTLPMTYTTATLNLILHOGY-IQFAERWKSVMY-----D 261
Qy 314 SSFM---KSIWATQLRDLATVYTLRYRQNPFCPSRNRRTAVSEFMKNTVHLIRNET-- 368
Db 262 ESFINQTKVDLQRRIQDYSTTVSTTFE-KFKPTLNPS-NKESVKNYR-----YVRSMTLQ 315
Qy 369 -----PYTIYGLDMSLSLYNETMFV-----ENKTASDNKTTTPTSPMG 408
Db 316 SLDIAATWPTLDNVNPSNVDIQLDQTRLVFSVAGPWEGNDNITSNIIDVLTPTINTGIG 375
Qy 409 FQ-----RTFIDPLWDYLDLSLLFLDEIRNFSLSRPTVYNLTTPPEH 448
Db 376 FQESSDLRKFTYPRIE-LQSMQFHGQ-----YVNSKSVEH 409

RESULT 10
AAW10653
ID AAW10653 standard; protein; 1167 AA.
XX
XX AC AAW10653;
XX
XX 25-MAR-2003 (revised)
XX 22-APR-1997 (first entry)
XX
XX Bacillus thuringiensis PS167P nematode active toxin.
XX
XX Nematode; active toxin; biological control; animal feed; additive;
XX production; recombinant; resistant; transgenic plant
XX
XX Bacillus thuringiensis.
XX
XX OS US5589382-A.
XX
XX PN
XX
XX ID AAW31504 standard; protein; 1167 AA.

```


Db 152 --PEGRGVGVAIYNKSNPKGAFIVLAPSGVG-----LLRYKRANDTYENASFPSR 200
 Qy 172 ---TFENPMFNPRWNTKL-----YV-GPTKN-----VDSOTIYFL 205
 Db 201 SKTTAMTWYNNFNSDELGGLLPROGNSNFFSARYVDSGDKVHSLTLTISSYSAYAI 260
 Qy 206 G-LTALLRYAQRNCTHSPYLVNAMSRLFRVPKYINGTKLKNTRKLR-----KQAPVK 260
 Db 261 GPLEA--GEYAWRD--DSKFYAVIRTSFGSFAFSKTLSSMDNVNIKALKACDVIPTSTN 317
 Qy 261 EQFEKAKKTOSTTT-----PYFS---YTTSAALN-----VTTNVT 293
 Db 318 IOFPTQIAKNYATPKLADNLASISVNCYPANKNIYLTSPFNVLVSGSETGMELSTST 377
 Qy 294 YSITT-----AARRVSTSTIAYRPDSSPMKSIM 321
 Db 378 ENITTLPPYVASLKSQSNICQANQADALRLVGSNKLAQQNMKSFQNIL 427

RESULT 13
 ID AAW22721 standard; protein; 1062 AA.
 AC AAW22721;

27-FEB-1998 (first entry)
 Antigenic protein derived from Mycoplasma gallisepticum.
 Mycoplasma gallisepticum antigen; epitope; monoclonal antibody; bird;
 mycoplasma infection; viral vaccine; avipoxvirus; herpesvirus.

Mycoplasma gallisepticum.

W09724370-A1.

10-JUL-1997.

27-DEC-1996; 96WO-JP003863.

28-DEC-1995; 95JP-00352754.

(JAPG) NIPPON ZEON KK.

Fujisawa A, Yoshida S;

WPI; 1997-363621/33.

N-PSDB; AAT75087.

Antigenic protein derived from Mycoplasma gallisepticum - useful in vaccines against, and for diagnosis of mycoplasma infection in birds.

Claim 3; Page 24-31; 45pp; Japanese.

This sequence represents an antigenic protein of Mycoplasma gallisepticum having an epitope specifically recognised by the monoclonal antibody (Mab) 35A6. This antigenic protein inhibits the metabolism of bird infecting mycoplasma. The antigenic protein and the DNA molecule can be used in the preparation of component and viral vaccines against mycoplasma infection in birds. Mycoplasma infection can be diagnosed by hybridising DNA in a sample with a DNA probe corresponding to 10 or more (preferably 15 to 40) bases of the antigenic protein gene

Sequence 1062 AA;

Query Match 4.2%; Score 103.5; DB 2; Length 1062;
 Best Local Similarity 18.9%; Pred. No. 5.3;
 Matches 98; Conservative 55; Mismatches 199; Indels 167; Gaps 22;

Qy 68 TTSTKVNVSQPVNLTMTNTEFPQYVYLACIQNSITYLWFDYFSTQLRKKPKYVVSQ 127
 Db 177 TDDGTFKFTKQ-----TQGEIVNDFILDAPILPKDLHPDWNLYIQRNILP----- 223

Qy 128 YNHTAKTITFRPPCGRVPSMTCISEMLNVSKRNDTSGQCGNFTTF-----NPMFFN--- 180
 Db 224 -NDVNTAVV--PWPVGRVSGTNADDMGDFGNGQITNDPIAQTAKTTTNDQNPFSTNSGA 280
 Qy 181 -----VPRWN-----TKLYVGPVKVNDVSQIY 203
 Db 281 MPGANNRYDSQLNVKHKRIKTSFOLDKFXVYPEWGTSEENKNITRLATGSLPSNERYWILD 340
 Qy 204 FLGLTALLRVAQRNCTHSPYLVNAMSRLFRVPKYING-----TKLKNTRM 250
 Db 341 IPGTPQVTLKEDSVNFSRLVNSVLSFGDSIYIFGTSELPSLWYSFPTRLSDLTA 400
 Qy 251 KLKKQAPVKEQBEKKAQKQSTTTPYPSYTTSAALNVTTNVTYSITTAARRVSTSI-A 309
 Db 401 LNQVKTDDEASSTDNGTTNGTTT-----TTDTSSGSGTAGTGTNTTSTQVSNPTLNT 455
 Qy 310 YRPDSSPMKSIMATQLRDLATW-----VYTLRYR-QNPFCEPRNRNTAVSEPMKN 359
 Db 456 YRSEGIDSKPTSANKI--DETWNADPNVIEARIYAYRLGIQNEI--PITN---AGNFIRN 509
 Qy 360 T-----HVLIR-----NETPYT-IVGTLD-----MSSLYVNETMFTVEN 391
 Db 510 TIGGVGTSTGSRVVLASVNGDORPTGNFQFLYVFGYLGYYQQTGTGTFWGYKLLNN 569
 Qy 392 -----KTASDSN---KTTPTSPSMG-----FQ 410
 Db 570 SPYDVLDSPRVGTETNQFRRTSLTPVNVGGVLTTEGARSFSTNTPYRAQGDTPESRSIFQ 629
 Qy 411 RTFDPLWDYLDLSLLFLDEIR---NFSLSRSTYNLTTP 446
 Db 630 SGYSDNTYEQSVLQVDFGIRNLLNVGVKASSFLNSNR 668

RESULT 14

ABR57376
 ID ABR57376 standard; protein; 1042 AA.

AC ABR57376;

11-SEP-2003 (first entry)

M. gallisepticum modified MGC3 antigen (M11-BTR) SEQ ID NO:78.

DNA molecule; prokaryotic cell; eukaryotic cell; virucide; vaccine; immunostimulant; viral infection.

Mycoplasma gallisepticum.

EPI275716-A2.

15-JAN-2003.

11-JUL-2002; 2002EP-00254879.

11-JUL-2001; 2001US-00901572.

25-APR-2002; 2002US-00131591.

(JAPG) ZEON CORP.

Okuda T, Saito S, Dorsey KM, Tsuzaki Y;

WPI; 2003-373746/36.

N-PSDB; ACP03435.

DNA molecule derived from a prokaryotic cell, useful for producing a vaccine for treating viral infections comprises at least one modified DNA regions encoding NXB so that no N-glycosylation occurs during expression.

Example 2; Page 59-62; 70pp; English.

The present invention describes a DNA molecule derived from a prokaryotic cell, where at least one of the DNA regions encoding NXB (where N =

```
CC asparagine, X = any amino acid other than proline, and B = serine or
CC threonine) has been modified so that no N-glycosylation occurs during the
CC expression in a eukaryotic cell. Also described: (1) a fused DNA
CC molecule, where a DNA encoding a signal sequence has been ligated to the
CC N-terminal end of the modified DNA molecule as described above so that it
CC may be expressed as a fusion protein; (2) a recombinant virus integrated
CC with the DNA molecule or the fused DNA molecule described above; (3)
CC producing a modified or fusion protein by using the recombinant virus
CC described above, to express a protein encoded by the modified DNA
CC molecule or the fused DNA molecule in a eukaryotic cell; and (4) a
CC vaccine comprising the recombinant virus. The DNA molecule has virucide
CC and immunostimulant activities. The DNA molecule is useful for producing
CC a vaccine for treating viral infections. The present sequence is used in
CC the exemplification of the present invention
XX
SQ Sequence 1042 AA;
Query Match 4.1%; Score 101.5; DB 6; Length 1042;
Best Local Similarity 18.1%; Pred. No. 7.9;
Matches 114; Conservative 73; Mismatches 223; Indels 221; Gaps 28;
QY 7 MVRDVKMFVLISIFLIVSPINCKVMSKALYNRPWRLGLVLS-KIGKYKLDQLKLE---- 61
Db 48 VLRDVDNFIITAQGTII-----KLD$---FTRPLVGLDLSDDFAGYKVKQIVSDYTT$ 98
QY 62 -----ILROLETTISTK-YNVSKQPVKNLTWTEP-----POY----- 93
Db 99 RNRFDQRTRAYALLVDEANVHLKRTNTNSNRGNQNSKRVIGGVNDPAHVIRPTD 159
QY 94 -----YILAGPIQNYSIITLVFDFEYSTQLRPAKYVSYQVNYHTAKTI 135
Db 159 DGTKEFQFTQGTGEIWNDFILDAPILPKDLHPDWNLYIQRLP-----NDVNTAV 210
QY 136 TPRPPPCGRVPSEMTCLSEMLNVSKENDTGEQCGGNFTTF-----NPMFFN 180
Db 211 V--PWPVGRVSGTNADGDFGNGQITNDPDLAQTKTITDNQNPSTFNSGAMPGANNRY 268
QY 181 -----VPRWN-----TKLYVGPTKVNDVSQTYIFLGLTALL 211
Db 269 DSQLVNKHRIKTSFQDKEFVPEWTEGSEENKQITRLATGSLPSNERYWIIDIPGTPOVT 328
QY 212 LRYAQNRNCTHSFVNVNMSRNLFRVPKYING-----TKLXNTWRKLRKQAP 258
Db 329 LKEDSVNVSRLYLVNSLSFGDSIYIFGTSPLSLWYISFPRLSDLTALNOVKTDD 388
QY 259 VKEQFEKAKKTQSTTPYFSTTGAALNVTNTVYSITTAARRYSTSTI-AYRPDSSFM 317
Db 389 IEASSTDQGTQTQGTIT-----TTDTSGSTGAGTGQTQTSTQTSVNSPNTLNTYR$FGIDS 443
QY 318 KSIMATQLDPLATW-----VYTLRYR-QNPFCEPSRNRRTAVSBFMKNT----- 360
Db 444 KPT$ANKI--DET$NADPNVIEARIYAEYRLGLQNEI--PITN---AGNFIRNTIGVGFT 497
QY 361 ----HVLIR-----NETPYT-IYGTILD-----MSLYNYNETMFEV----- 391
Db 498 STGSRVLRASNGDQRPNGFQFPFLYVFGYQQTGTGTWYGYTKLLQSPYDVLDA 557
QY 392 -KTASDN---KTTPTSPMG-----FRTTIDPLW 418
Db 558 ARVGTETNQFRRTSLTYPMVGGYLTBEGAR$F$NTPIYIRAQGTPTESRIFSGYSDNTY 617
QY 419 DYLD$LLFLDEIR---N$FLSRPTYVNLTPP 446
Db 618 EYIQ$VLGFDGIRN$NLNVGVK$SFLNSRNP 648
RESULT 15
ID ADO28930
XX ADO28930 standard; protein; 2481 AA.
AC ADO28930;
XX
DT 29-JUL-2004 (first entry)
```

```
XX
DE Human novel GPCR PGRI7, SEQ ID NO:29.
XX
XX G protein-coupled receptor; GPCR; drug screening; diagnosis;
XX transgenic mouse; neurological disorder; adrenal gland disorder;
XX colon disorder; intestinal disorder; cardiovascular disorder;
XX muscular disorder; blood disorder; immune disorder; bone disorder;
XX joint disorder; metabolic disorder; nutritive disorder; cancer;
XX kidney disorder; liver disorder; lung disorder; breast disorder;
XX ovary disorder; uterus disorder; prostate disorder; spleen disorder;
XX skin disorder; stomach disorder; antiparkinsonian; antiemetic;
XX muscle disorder; thyroid disorder; vasotropic; antiarrhythmic;
XX cytostatic; antiinflammatory; respiratory; antidiarrhoeic; antidiabetic;
XX CNS; central nervous system; antibacterial; antianaemic; antisporrheic;
XX virucide; hepatotropic; antitubercular; antiallergic; anorectic;
XX dermatological; antiulcer; antithyroid; antihypertensive; anorectic;
XX immunosuppressive; nephrotropic; gene therapy; GPCR modulator; human;
XX receptor.
XX
XX Homo sapiens.
OS
XX
XX WO2004040000-A2.
PN
XX
XX 13-MAY-2004.
PD
XX
XX 09-SEP-2003; 2003WO-US028226.
PF
XX
XX 09-SEP-2002; 2002US-0409303P.
PR
XX
XX 09-APR-2003; 2003US-0461329P.
PA
XX
XX (PRIM-) PRIMAL INC.
PI
XX
XX Gaitanaris GA, Bergmann JE, Gragerov A, Hohmann J, Li F;
XX Madisen L, McIlwain KL, Pavlova MN, Vassiliadis D, Zeng H;
XX
XX WPI; 2004-390329/36.
DR
XX
XX N-PSDB; ADO28931.
XX
XX Novel mammalian G protein coupled receptors, useful for identifying
XX compounds that modulates diagnosing and treating disease condition
XX associated with GPCR dysfunction e.g. autoimmune diseases, angina
XX pectoris, Parkinson's disease.
XX
XX Claim 1; SEQ ID NO 29; 542pp; English.
XX
XX The invention relates to human and mouse G protein-coupled receptors
XX (GPCRs) and nucleic acids encoding them. The invention also relates to
XX sequences at least 90% identical to the GPCR proteins and nucleic acids
XX of the invention; methods of treating, preventing or diagnosing diseases
XX associated with GPCRs of the invention; methods of screening for
XX compounds useful in the treatment of GPCR-related diseases; a transgenic
XX mouse comprising a GPCR gene of the invention; a mouse comprising a
XX mutation in a GPCR transgene or in an endogenous GPCR gene; cells derived
XX from the transgenic mice; kits comprising several mice, each of which has
XX a mutation in a different GPCR gene of the invention; and kits comprising
XX probes which hybridise to GPCR polynucleotides of the invention. The
XX invention further discloses variants of the GPCR polypeptides and vectors
XX comprising a GPCR nucleic acid. The GPCR nucleic acids and proteins may
XX be used in the diagnosis, treatment or prevention of a wide variety of
XX diseases including neurological disorders (e.g., Alzheimer's disease,
XX depression, diabetic neuropathy, Parkinson's disease or schizophrenia);
XX disorders of the adrenal gland; disorders of the colon or intestine
XX (e.g., Crohn's disease, diarrhoea, food poisoning or irritable bowel
XX syndrome); cardiovascular disorders (e.g., angina, cardiac arrhythmia or
XX myocardial infarction); muscular disorders; blood disorders (e.g.,
XX anaemia or leukaemia); immune disorders (e.g., osteoarthritis, rheumatoid
XX AIDS); bone and joint disorders (e.g., osteoporosis); metabolic or nutritive disorders (e.g.,
XX arthritis, gout or osteoporosis); metabolic or nutritive deficiency-related
XX diseases); and disorders of the kidney, liver, lung, breast, ovary,
XX uterus, prostate, testis, skin, stomach, pancreas, spleen, thymus and
XX thyroid (e.g., cancers). The present sequence represents a GPCR of the
XX invention. Note: The full sequence data for this patent did not form part
```


CC of the printed specification; those sequences not shown were obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2481 AA;

Query Match 4.1%; Score 101; DB 8; Length 2481;
 Best Local Similarity 18.8%; Pred. No. 30; Mismatches 150; Indels 140; Gaps 19;
 Matches 81; Conservative 61; Mismatches 150; Indels 140; Gaps 19;
 Qy 69 TISTKYNVSKQFVNKLNTNTEFPQYIILAGPIQNYIITYLWDF-----112
 Db 1757 SLHTSFNIQVSP--SLT-----SFKSASGPKKNVKTITNCFSNTRKWTSLLEKTSLTN 1808
 Qy 113 YSTQLRKPQY-----VYQYNHTAKTITFR--PPCGRVPSMTCLSEM 154
 Db 1809 YATSLNTPVSPYPPWTPSSATLPSLTSFYVSPHSTEABISTPKISPPPTSQMVPEPVLGTR 1868
 Qy 155 LNVSKRNDTGCGCCNFTFNPMPFNPRWTKLYVG-----PTKVNVDSTIY 203
 Db 1869 M-----TSSNTQPLLMTSNIPITAECSQFPISITINVPSTNEMETETLH 1912
 Qy 204 FL-----GLTALLLR--YAQRNCTHSFVLVN-AMSRNLFVRPKYINGTKLKNTRK 251
 Db 1913 LVPGLSTFTTASQGLYSKDYMMAMSSIPMSGILPNHGLSEN-----PSLSTSLRA 1962
 Qy 252 LKRQAPVKEQFEKAKK-TOSTTTPYF--SYTTSALNVTNTVYSITTA-----299
 Db 1963 ITSTLADVKHTEFKWTSVTPGTTLPSTLGSATSGSVISKSPILTWLLSSLPSPGPATV 2022
 Qy 300 --ARRVSTTIAYRPDSFMKSIATQRLDIAWYITLYRQNPFPCEPSNRRTAVSEFM 357
 Db 2023 SNAPHVMTSTSTVEVSKSTFLTSDMIS-----AHPFTNLTLPSAT 2062
 Qy 358 KNTHLIRNETPYTIYG-----TLDMSLYYNETMFV-----ENKTASDSNKTPTS 404
 Db 2063 MST-ILTRTIPTPLGGITGTPPSLPSMNVTDIVITSTHPEASSKTTITANPRVSH 2121
 Qy 405 PSMGFORTPIDP 416
 Db 2122 PS-SFSRKTWSP 2132

Search completed: December 7, 2004, 15:17:13
 Job time : 160 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:03:50 ; Search time 200 Seconds
(without alignments)
1337.746 Million cell updates/sec

Title: US-09-942-146A-1
Perfect score: 2451
Sequence: 1 MGRKEMVRDVKMFLVLSI.....PEHRRVNLSTNSLWMLQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	2416.5	98.6	466	1	UL74_HCMVA	P16750 human cytom
2	2416.5	98.6	466	2	Q7M6L8	Q7m6l8 human cytom
3	2416.5	98.6	466	2	AAN40055	Aan40055 human cyt
4	2410.5	98.3	466	2	Q8AZ28	Q8az28 human cytom
5	2406.5	98.2	466	2	Q8BCU8	Q8bcu8 human cytom
6	2397.5	97.8	466	2	Q8AZ36	Q8az36 human cytom
7	2397.5	97.8	466	2	Q8BCU2	Q8bcu2 human cytom
8	2068	84.4	463	2	Q8AZ32	Q8az32 human cytom
9	2065	84.3	463	2	Q8BCV2	Q8bcv2 human cytom
10	2065	84.3	463	2	Q8AZ39	Q8az39 human cytom
11	2064	84.2	463	2	Q8BCU3	Q8bcu3 human cytom
12	2063.5	84.2	464	2	Q8BCU8	Q8bcu8 human cytom
13	2051.5	83.7	464	2	Q8BCV2	Q8bcv2 human cytom
14	2051.5	83.7	464	2	Q8BCU2	Q8bcu2 human cytom
15	2050.5	83.7	462	2	Q8BCU0	Q8bcu0 human cytom
16	2045.5	83.5	462	2	Q8BCU6	Q8bcu6 human cytom
17	1942	79.2	465	2	Q8AYW0	Q8ayw0 human cytom
18	1936	79.0	465	2	Q8BCU9	Q8bcu9 human cytom
19	1899	77.5	462	2	Q8AYZ3	Q8ayz3 human cytom
20	1894.5	77.3	462	2	Q8BCV5	Q8bcv5 human cytom
21	1870.5	76.3	472	2	Q8BCV1	Q8bcv1 human cytom
22	1870.5	76.3	472	2	AAR31626	Aar31626 human cyt
23	1748.5	71.3	464	2	Q8BCU1	Q8bcu1 human cytom
24	1740.5	71.0	464	2	Q8BCU4	Q8bcu4 human cytom
25	1737.5	70.9	464	2	Q8AYV5	Q8ayv5 human cytom
26	1735.5	70.8	464	2	Q8AYZ5	Q8ayz5 human cytom
27	1730	70.6	457	2	Q8BCV7	Q8bcv7 human cytom
28	1730	70.6	457	2	Q8BCV0	Q8bcv0 human cytom
29	977.5	39.9	461	2	Q8QS25	Q8qs25 chimpanzee
30	772	31.5	389	2	Q7TFN1	Q7tfn1 rhesus cyto
31	717.5	29.3	144	2	Q7T9B3	Q7t9b3 human cytom

32	717.5	29.3	144	2	Q7T9B4	Q7t9b4 human cytom
33	717.5	29.3	144	2	Q7T9B5	Q7t9b5 human cytom
34	717.5	29.3	144	2	Q7T9B6	Q7t9b6 human cytom
35	717.5	29.3	144	2	Q7T9C3	Q7t9c3 human cytom
36	717.5	29.3	144	2	Q7T9C4	Q7t9c4 human cytom
37	717.5	29.3	144	2	Q7T9C5	Q7t9c5 human cytom
38	717.5	29.3	144	2	Q7T9D2	Q7t9d2 human cytom
39	717.5	29.3	144	2	Q7T9D4	Q7t9d4 human cytom
40	717.5	29.3	144	2	Q7T9D5	Q7t9d5 human cytom
41	706.5	28.8	144	2	Q7T9B2	Q7t9b2 human cytom
42	678.5	27.7	144	2	Q7T9C1	Q7t9c1 human cytom
43	497	20.3	144	2	Q7T9B8	Q7t9b8 human cytom
44	497	20.3	144	2	Q7T9D0	Q7t9d0 human cytom
45	491	20.0	142	2	Q7T9B9	Q7t9b9 human cytom

ALIGNMENTS

RESULT 1						
ID	UL74_HCMVA	STANDARD;	PRT;	466	AA.	
AC	P16750;					
DT	01-AUG-1990 (Rel. 15, Created)					
DT	01-AUG-1990 (Rel. 15, Last sequence update)					
DT	05-JUL-2004 (Rel. 44, Last annotation update)					
DE	Glycoprotein-UL74-precursor;					
GN	Name=UL74;					
OS	Human cytomegalovirus (strain AD169).					
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;					
OC	Betaherpesvirinae; Cytomegalovirus.					
OX	NCBI_TaxID=10360;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE:90269038; PubMed:2161319;					
RA	Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,					
RA	Horsnell T., Hutchison C.A. III, Kourazides T., Martignetti J.A.,					
RA	Freddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;					
RT	"Analysis of the protein-coding content of the sequence of human					
RT	cytomegalovirus strain AD169.";					
RT	Curr. Top. Microbiol. Immunol. 154:125-169(1990).					
CC	-1- SIMILARITY: Belongs to the herpesviruses/UL7 family					

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).					
CC	-----					
CC	EMBL: X17403; CAA35389.1; -;					
DR	PIR; S09837; S09837;					
KW	Glycoprotein; Signal.					
FT	CHAIN	1	30			Potential.
FT	31	466				Glycoprotein UL74.
FT	CARBOHYD	75	75			N-linked (GlcNAc. .)
FT	CARBOHYD	83	83			N-linked (GlcNAc. .)
FT	CARBOHYD	87	87			N-linked (GlcNAc. .)
FT	CARBOHYD	103	103			N-linked (GlcNAc. .)
FT	CARBOHYD	130	130			N-linked (GlcNAc. .)
FT	CARBOHYD	157	157			N-linked (GlcNAc. .)
FT	CARBOHYD	162	162			N-linked (GlcNAc. .)
FT	CARBOHYD	171	171			N-linked (GlcNAc. .)
FT	CARBOHYD	219	219			N-linked (GlcNAc. .)
FT	CARBOHYD	242	242			N-linked (GlcNAc. .)
FT	CARBOHYD	288	288			N-linked (GlcNAc. .)
FT	CARBOHYD	292	292			N-linked (GlcNAc. .)
FT	CARBOHYD	350	350			N-linked (GlcNAc. .)
FT	CARBOHYD	385	385			N-linked (GlcNAc. .)
FT	CARBOHYD	392	392			N-linked (GlcNAc. .)
FT	CARBOHYD	399	399			N-linked (GlcNAc. .)
FT	CARBOHYD	433	433			N-linked (GlcNAc. .)

Query Match 454 454 N-linked (GlcNAc...) (Potential).
Best Local Similarity 98.6%; Score 2416.5; DB 1; Length 466;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Db 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60

QY 61 EILROLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 119
Db 61 EILROLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 120

QY 120 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

QY 180 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 240

QY 240 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300

QY 300 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 359
Db 301 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 360

QY 360 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 419
Db 361 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 420

QY 420 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 465
Db 421 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 466

RESULT 2
Q7M6L8 PRELIMINARY; PRT; 466 AA.
ID Q7M6L8
AC Q7M6L8
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD169;
RX MEDLINE=22421467; PubMed=12533697; Addison C., Dargan D.J.,
RA Davidson A.J., Dolan A., Akter P., Hayward G.S.;
RA Alexander D.J., McGeoch D.J., Hayward G.S.;
RT "The human cytomegalovirus genome revisited: comparison with the
RT chimpanzee cytomegalovirus genome.";
RL J. Gen. Virol. 84:17-28(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AD169;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The genes encoding the gCII complex of human cytomegalovirus exist
RT in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR ENBL; BK000394; DAA00170.1; -;
DR ENBL; AF531331; AAN40055.1; -;
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 454 454 N-linked (GlcNAc...) (Potential).
Best Local Similarity 98.9%; Score 2416.5; DB 2; Length 466;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Db 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60

QY 61 EILROLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 119
Db 61 EILROLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 120

QY 120 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

QY 180 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 240

QY 240 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300

QY 300 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 359
Db 301 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 360

QY 360 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 419
Db 361 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 420

QY 420 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 465
Db 421 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 466

RESULT 3
AAN40055 PRELIMINARY; PRT; 466 AA.
ID AAN40055
AC AAN40055
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD169;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The Genes Encoding the gCII Complex of Human Cytomegalovirus Exist
RT in Highly Diverse Combinations in Clinical Isolates.";
RL J. Virol. 76:10841-10848(2002).
DR ENBL; AF531331; AAN40055.1; -;
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 454 454 N-linked (GlcNAc...) (Potential).
Best Local Similarity 98.9%; Score 2416.5; DB 2; Length 466;
Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60
Db 1 MGRKEMVDRVDPKMFVLISIFLLVSFINCKVMSKALYNRPWRGLVLSKIGKYKLDQLKL 60

QY 61 EILROLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 119
Db 61 EILROLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYSTIYLWFDYFYSTQLRK 120

QY 120 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSOYNHTAKITIFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

QY 180 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 239
Db 181 NVRWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPKY 240

QY 240 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQFEKAKKTQSTTTPYFSTTSAAALNVTNTVYSITTA 300

QY 300 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 359
Db 301 ARVSTSTIAYRPDSSFMKSIMATQRLDRLATVVTTLRYQNPFCPEPSRNRATAVSEFMKN 360

QY 360 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 419
Db 361 THVLIRNETPTTYTGTLDMSSLYNNTFVENKTASDSNKTTPSPSGMFGQRTFIDPLWD 420

QY 420 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 465
Db 421 YLDSLLFLDEIRNFSRPTVYNLTTPPEHRAVNLSTNSLWWLQ 466

RESULT 3
AAN40055 PRELIMINARY; PRT; 466 AA.
ID AAN40055
AC AAN40055
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD169;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
RA "The Genes Encoding the gCII Complex of Human Cytomegalovirus Exist
RT in Highly Diverse Combinations in Clinical Isolates.";
RL J. Virol. 76:10841-10848(2002).
DR ENBL; BK000394; DAA00170.1; -;
DR ENBL; AF531331; AAN40055.1; -;
SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

```
Db 121 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 239
Db 181 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 240
Qy 240 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPYIYGTLDMSLLYNETMFVENKTASDSNKTTPTSPSGMGTFTFDPLWD 419
Db 420 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 466
```

RESULT 4

```
Q8AZ28
ID Q8AZ28 PRELIMINARY; PRT; 466 AA.
AC Q8AZ28;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=298, SW1, and SW3;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531320; AAN40044.1; -
DR EMBL; AF531325; AAN40049.1; -
DR EMBL; AF531337; AAN40061.1; -
DR EMBL; AF531346; AAN40070.1; -
SQ SEQUENCE 466 AA; 54200 MW; 56581B75919572FE CRC64;
```

```
Query Match 98.3%; Score 2410.5; DB 2; Length 466;
Best Local Similarity 98.7%; Pred. No. 1.3e-177;
Matches 460; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MGRKEMVDRVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIKYKLDQKL 60
Db 1 MGRKEMVDRVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIKYKLDQKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEPPQYIILAGPIQNTSYTLWDFDYSTQLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTWNTEPPQYIILAGPIQNTSYTLWDFDYSTQLRK 120
Qy 120 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 239
Db 181 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 240
Qy 240 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 EILRQLETTISTKYNVSKQPVKNLTWNTEPPQYIILAGPIQNTSYTLWDFDYSTQLRK 120
Qy 120 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 239
Db 181 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 240
Qy 240 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
```

```
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPYIYGTLDMSLLYNETMFVENKTASDSNKTTPTSPSGMGTFTFDPLWD 419
Db 361 THVLIRNETPYIYGTLDMSLLYNETMFVENKTASDSNKTTPTSPSGMGTFTFDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 466
```

RESULT 5

```
Q8BCU8
ID Q8BCU8 PRELIMINARY; PRT; 466 AA.
AC Q8BCU8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=851;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531330; AAN40054.1; -
SQ SEQUENCE 466 AA; 54230 MW; 56581B75919562FF CRC64;
```

```
Query Match 98.2%; Score 2406.5; DB 2; Length 466;
Best Local Similarity 98.5%; Pred. No. 2.6e-177;
Matches 459; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 MGRKEMVDRVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIKYKLDQKL 60
Db 1 MGRKEMVDRVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIKYKLDQKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEPPQYIILAGPIQNTSYTLWDFDYSTQLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTWNTEPPQYIILAGPIQNTSYTLWDFDYSTQLRK 120
Qy 120 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYISQYNHTAKTITFRPPCGVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 239
Db 181 NVRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFYLNVAMSRNLFVRPKY 240
Qy 240 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSI MATQLDLATWVYTLRYQRNPPCEPSRNRNAVSEFMKN 360
Qy 360 THVLIRNETPYIYGTLDMSLLYNETMFVENKTASDSNKTTPTSPSGMGTFTFDPLWD 419
Db 361 THVLIRNETPYIYGTLDMSLLYNETMFVENKTASDSNKTTPTSPSGMGTFTFDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLRSPTYVNLTPPEHRAVNLSNSTNSLWMLQ 466
```

```
RESULT 6
Q8AZ36 Q8AZ36 PRELIMINARY; PRT; 466 AA.
AC Q8AZ36;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC855844E3B65A0 CRC64;

Query Match 97.8%; Score 2397.5; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 1.3e-176;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKMMVRDVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
DB 1 MGRKMMVRDVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNTSYIYLFDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNTSYIYLFDFYSTQLRK 120
QY 120 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVRPWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
DB 181 NVRPWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
QY 240 INGTKLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTKLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIATQLRDLATWVYTLRYQNPFCPEPSRNRATVSEPMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIATQLRDLATWVYTLRYQNPFCPEPSRNRATVSEPMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 7
Q8BCU2 Q8BCU2 PRELIMINARY; PRT; 466 AA.
AC Q8BCU2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;

Query Match 97.8%; Score 2397.5; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 1.3e-176;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKMMVRDVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
DB 1 MGRKMMVRDVPKMFVLISISFLVSNCKVMSKALYNRPWRGLVLSKIGYKLDQLK 60
QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNTSYIYLFDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNTSYIYLFDFYSTQLRK 120
QY 120 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVSYQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVRPWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
DB 181 NVRPWNTKLYVGPVKVNDSTQIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
QY 240 INGTKLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 299
DB 241 INGTKLKNTMKLRKQAPVKEQFEKAKKTQSTITTPYFSYTTSAALNVTTNVTYSITTA 300
QY 300 ARRVSTSTIAYRPDSSFMKSIATQLRDLATWVYTLRYQNPFCPEPSRNRATVSEPMKN 359
DB 301 ARRVSTSTIAYRPDSSFMKSIATQLRDLATWVYTLRYQNPFCPEPSRNRATVSEPMKN 360
QY 360 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLWD 419
DB 361 THVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLRSPYVNLTPPEHRRVAVNLSTNSLWMLQ 466

RESULT 8
Q8AZ32 Q8AZ32 PRELIMINARY; PRT; 463 AA.
AC Q8AZ32;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;
```

```
Query Match      84.4%; Score 2068; DB 2; Length 463;
Best Local Similarity 85.4%; Pred. No. 3.5e-151;
Matches 399; Conservative 22; Mismatches 40; Indels 6; Gaps 4;

Qy 1 MGRKEMVVDVPMKMFVLISISFLVSPFNCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEM-RGVNLFPLMSLTLFLFSFNCRAAVLSVGRYWSGKVLSTIGKQRLDKPKL 59

Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNSYIYLFDFYSTQLR 118
Db 60 EILQLEKDIYTKYFNMTROHIKNLTWNTEFPYIILAGPIQNSVYLYLWDFYSTQLR 119

Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLEMLNVSKRNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEYNHTAKTITFRPPSCGTVPMSCTLEMLNVSKRNDTGEQCGNFTTFNPMF 179

Qy 179 FNVPRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRWNTKLVGSKKVNVDSTQIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239

Qy 239 YINGTKLNTWRKLRKQAPVKEQEKAKTKTQSTTTTTFYFYTSSAALNVTNTVYSITT 298
Db 240 YINGTKLNTWRKLRKQAPVKEQEKSKKSQSTTTTTFYFYTSSAALNVTNTVYSITT 299

Qy 299 AARVSTSTIAYRPDSSPFMSIMATQRLDRLATWVYTLRYQNPFCPSRNRRTAVSEPMK 358
Db 300 TARRVSTSTIAYRPDSSPFMSIMTTLQRLDRLATWVYTLRYQNPFCPSRNRRTAVSEPMK 359

Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTTTPSPMGFORFIDPLW 418
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGFORFIDPLW 416

Qy 419 DYLDLFLDLIRNFSLSRSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSLSQSPYGNLTPEHRRAVNLSTNSLWMLQ 463

RESULT 9
Q8BCV9 PRELIMINARY; PRT; 463 AA.
AC Q8BCV9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SW990;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531354; AAN40078.1; -.
SQ SEQUENCE 463 AA; 53885 MW; E86783A05FB9424A CRC64;

Query Match      84.3%; Score 2065; DB 2; Length 463;
Best Local Similarity 85.2%; Pred. No. 5.9e-151;
Matches 398; Conservative 23; Mismatches 40; Indels 6; Gaps 4;

Qy 1 MGRKEMVVDVPMKMFVLISISFLVSPFNCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEM-RGVNLFPLMSLTLFLFSFNCRAAVLSVGRYWSGKVLSTIGKQRLDKPKL 59

Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNSYIYLFDFYSTQLR 118
Db 60 EILQLEKDIYTKYFNMTROHIKNLTWNTEFPYIILAGPIQNSVYLYLWDFYSTQLR 119

Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLEMLNVSKRNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEYNHTAKTITFRPPSCGTVPMSCTLEMLNVSKRNDTGEQCGNFTTFNPMF 179

Qy 179 FNVPRWNTKLVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRWNTKLVGSKKVNVDSTQIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239

Qy 239 YINGTKLNTWRKLRKQAPVKEQEKAKTKTQSTTTTTFYFYTSSAALNVTNTVYSITT 298
Db 240 YINGTKLNTWRKLRKQAPVKEQEKSKKSQSTTTTTFYFYTSSAALNVTNTVYSITT 299

Qy 299 AARVSTSTIAYRPDSSPFMSIMATQRLDRLATWVYTLRYQNPFCPSRNRRTAVSEPMK 358
Db 300 TARRVSTSTIAYRPDSSPFMSIMTTLQRLDRLATWVYTLRYQNPFCPSRNRRTAVSEPMK 359

Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTTTTPSPMGFORFIDPLW 418
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTPTSPSTGFORFIDPLW 416

Qy 419 DYLDLFLDLIRNFSLSRSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSLSQSPYGNLTPEHRRAVNLSTNSLWMLQ 463

RESULT 9
Q8BCV9 PRELIMINARY; PRT; 463 AA.
AC Q8BCV9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SW990;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531354; AAN40078.1; -.
SQ SEQUENCE 463 AA; 53885 MW; E86783A05FB9424A CRC64;

Query Match      84.3%; Score 2065; DB 2; Length 463;
Best Local Similarity 85.2%; Pred. No. 5.9e-151;
Matches 398; Conservative 23; Mismatches 40; Indels 6; Gaps 4;

Qy 1 MGRKEMVVDVPMKMFVLISISFLVSPFNCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEM-RGVNLFPLMSLTLFLFSFNCRAAVLSVGRYWSGKVLSTIGKQRLDKPKL 59

Qy 61 EILQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNSYIYLFDFYSTQLR 118
Db 60 EILQLEKDIYTKYFNMTROHIKNLTWNTEFPYIILAGPIQNSVYLYLWDFYSTQLR 119

Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLEMLNVSKRNDTGEQCGNFTTFNPMF 178
```


Db 300 TARRVSTSTIAYRPDSSFMKSIIMTQLRDLATWVYVTLRYQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTPTTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTTSPSTGQRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 463

RESULT 11
Q8AZ39
ID Q8AZ39 PRELIMINARY; PRT; 463 AA.
AC Q8AZ39
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=650, and DM7;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848 (2002).
DR EMBL; AF531326; AAN40050.1; -
DR EMBL; AF531334; AAN40058.1; -
SQ SEQUENCE 463 AA; 53901 MW; 403100AABDF4412A CRC64;

Query Match 84.2%; Score 2064; DB 2; Length 463;
Best Local Similarity 85.2%; Pred. No. 7e-151;
Matches 398; Conservative 22; Mismatches 41; Indels 6; Gaps 4;
Qy 1 MGRKEMVVDVPMFVLISIFLVSFINKVMSKALNRPWRGLVLSKIGYKLDOLKL 60
Db 1 MGRKEM-RGVNLFPLSLTFLSFINKVMSKALNRPWRGLVLSKIGYKLDOLKL 59
Qy 61 EILRQLETTISTKY-NVSKQPVKMLTN-TEFPQYIILAGPIQNYSTYILWDFYSTQLR 118
Db 60 EILKQLEKDIYTKYFNMTQHKMLTNMTTEFPYIILAGPIQNSVYILWDFYSTQLR 119
Qy 119 KPAKYVSYQNHNTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 178
Db 120 KPAKYVSEYNHTAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGNFTFNPMP 179
Qy 179 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHGFYLVNMSRLFRVPK 238
Db 180 FNVPRWNTKLVGSKKNVDSTIYFLGLTALLRYAQRNCTHGFYLVNMSRLFRVPK 239
Qy 239 YINGTKLNTWTKLRKQAPVKEQFEKAKTKTQSTTTPYFSTYTTSAALNVTTNVTYSITT 298
Db 240 YINGTKLNTWTKLRKQAPVKEQSEKSKSQSTTTPYFSTYTTSTALNVTTNATYSVTT 299
Qy 299 AARVSTSTIAYRPDSSFMKSIIMATQRLDRLATWVYVTLRYQNPFCSSRNRTAVSEFMK 358
Db 300 TTRVSTSTIAYRPDSSFMKSIIMTQLRDLATWVYVTLRYQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTPTTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASE--TTTSPSTGQRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 463

RESULT 12

Q8BCU3
ID Q8BCU3 PRELIMINARY; PRT; 464 AA.
AC Q8BCU3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SW1324;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848 (2002).
DR EMBL; AF531340; AAN40064.1; -
SQ SEQUENCE 464 AA; 54146 MW; A24A513BF112848B CRC64;

Query Match 84.2%; Score 2063.5; DB 2; Length 464;
Best Local Similarity 84.8%; Pred. No. 7.7e-151;
Matches 396; Conservative 25; Mismatches 41; Indels 5; Gaps 4;
Qy 1 MGRKEMVVDVPMFVLISIFLVSFINKVMSKALNRPWRGLVLSKIGYKLDOLKL 60
Db 1 MGRKEDM-RSISKLFPIISITVLLFSINCKVVRPP--GRYWLGTVLSTIGKQKLDKPKL 57
Qy 61 EILRQLETTISTKY-NVSKQPVKMLTN-TEFPQYIILAGPIQNYSTYILWDFYSTQLR 118
Db 58 EILKQLEREPVTKYFNMTQHKMLTNMTQFPQYIILAGPIRNDSTYILWDFYSTQLR 117
Qy 119 KPAKYVSYQNHNTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMP 178
Db 118 KPAKYVSYQNHNTAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGNFTFNPMP 177
Qy 179 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHGFYLVNMSRLFRVPK 238
Db 178 FNVPRWNTKLVGPTKVNVDSTIYFLGLTALLRYAQRNCTHGFYLVNMSRLFRVPK 237
Qy 239 YINGTKLNTWTKLRKQAPVKEQFEKAKTKTQSTTTPYFSTYTTSAALNVTTNVTYSITT 298
Db 238 YINGTKLNTWTKLRKQAPVKEQLEKTKKSQSTTTPYFSTYTTSTALNVTTNATYRVT 297
Qy 299 AARVSTSTIAYRPDSSFMKSIIMATQRLDRLATWVYVTLRYQNPFCSSRNRTAVSEFMK 358
Db 298 SAKRIPTSTIAYRPDSSFMKSIIMATQRLDRLATWVYVTLRYRNEPFCPPDRNRTAVSEFMK 357
Qy 359 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTPTTSPSMGFQRTFIDPLW 418
Db 358 NTHVLIRNETPTTYIGTLDMSLYNETMFVENKTASDNKTPTTSPSTGQRTFIDPLW 417
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 465
Db 418 DYLDLFLDLIRNFSRLSPYVNLTPPEHRRVAVNLSTNSLWMLQ 464

RESULT 13
Q8BCT8
ID Q8BCT8 PRELIMINARY; PRT; 464 AA.
AC Q8BCT8
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE UL74 protein.
GN Name=UL74; ORFNames=HHV5gp069;
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=Toledo;
 RX MEDLINE=22255414; PubMed=12368327;
 RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist
 in highly diverse combinations in clinical isolates.";
 RL J. Virol. 76:10841-10848(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Toledo;
 RA Brondke H., Schmitz B., Shenk T., Doerfler W.;
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF531355; AAN40073.1; -;
 DR EMBL; AY486475; AAS48965.1; -;
 SQ SEQUENCE 464 AA; 54020 MW; 7751A6829CEB3E7 CRC64;
 Query Match 83.7%; Score 2051.5; DB 2; Length 464;
 Best Local Similarity 84.4%; Pred. No. 6.5e-150;
 Matches 394; Conservative 25; Mismatches 43; Indels 5; Gaps 4;
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGKYKLDQKL 60
 Db 1 MGRKGM-RSISKLFPIISLTVLLFSLINCKVVRPP--GRYMLGTVLSTIGKQKLDKFKL 57
 Qy 61 EILRQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118
 Db 58 EILKQLEREPTYKFNMTQHVKNLTWNMTQFPQYIILAGPIRNDSTIYLWFDYFYSTQLR 117
 Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 178
 Db 118 KPAKYVYSQYNHTAKTITFRPPSCGTVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 177
 Qy 179 FNVPRMNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPK 238
 Db 178 FNVPRMNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPK 237
 Qy 239 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 298
 Db 238 YINGTKLNTMRKLRKQAPVKEQLEKTKKSQSTTTTPYSYTTSTALNVTNTYIKVTT 297
 Qy 299 AARVSTSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYRONPFCPSRNRRTAVSEFMK 358
 Db 298 SAKRIPSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYRNEFCPKDRNRRTAVSEFMK 357
 Qy 359 NTHVLIRNETPYTYIGTLDSSLYNETMVENKTASDNKTTTSPSMGFQRTFIDPLW 418
 Db 358 NTHVLIRNETPYTYIGTLDSSLYNETMVENKTASDNKTTTSPSTRFQKTFIDPLW 417
 Qy 419 DYLDLSLLFLDEIRNFSRLSPYVNLTPPEHRAVNLSTSLMWLQ 465
 Db 418 DYLDLSLLFLDKIRNFSRLQPAYGNLTPEHRAVNLSTSLMWWSQ 464
 RESULT 14
 AAS48965 PRELIMINARY; PRT; 464 AA.
 AC AAS48965;
 DT 10-MAR-2004 (TrEMBLrel. 27, Created)
 DT 10-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 10-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE UL74.
 GN UL74 OR HHV5GP069.
 OS Human cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Toledo;
 RA Brondke H., Schmitz B., Shenk T., Doerfler W.;
 RL "Differences in the Nucleotide Sequences between the AD169 and Toledo
 Strains of Human Cytomegalovirus.";
 RT Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY486475; AAS48965.1; -;
 SQ SEQUENCE 464 AA; 54020 MW; 7751A6829CEB3E7 CRC64;
 Query Match 83.7%; Score 2051.5; DB 2; Length 464;
 Best Local Similarity 84.4%; Pred. No. 6.5e-150;
 Matches 394; Conservative 25; Mismatches 43; Indels 5; Gaps 4;
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGKYKLDQKL 60
 Db 1 MGRKGM-RSISKLFPIISLTVLLFSLINCKVVRPP--GRYMLGTVLSTIGKQKLDKFKL 57
 Qy 61 EILRQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118
 Db 58 EILKQLEREPTYKFNMTQHVKNLTWNMTQFPQYIILAGPIRNDSTIYLWFDYFYSTQLR 117
 Qy 119 KPAKYVYSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 178
 Db 118 KPAKYVYSQYNHTAKTITFRPPSCGTVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMF 177
 Qy 179 FNVPRMNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPK 238
 Db 178 FNVPRMNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHFSYLVNMSRNLFRVPK 237
 Qy 239 YINGTKLNTMRKLRKQAPVKEQFEKAKTKTQSTTTTPYSYTTSAALNVTNTYISITT 298
 Db 238 YINGTKLNTMRKLRKQAPVKEQLEKTKKSQSTTTTPYSYTTSTALNVTNTYIKVTT 297
 Qy 299 AARVSTSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYRONPFCPSRNRRTAVSEFMK 358
 Db 298 SAKRIPSTIAYRPDSSFMKSIMATQRLDWTATVYTLRYRNEFCPKDRNRRTAVSEFMK 357
 Qy 359 NTHVLIRNETPYTYIGTLDSSLYNETMVENKTASDNKTTTSPSMGFQRTFIDPLW 418
 Db 358 NTHVLIRNETPYTYIGTLDSSLYNETMVENKTASDNKTTTSPSTRFQKTFIDPLW 417
 Qy 419 DYLDLSLLFLDEIRNFSRLSPYVNLTPPEHRAVNLSTSLMWLQ 465
 Db 418 DYLDLSLLFLDKIRNFSRLQPAYGNLTPEHRAVNLSTSLMWWSQ 464
 RESULT 15
 Q8BCU0 PRELIMINARY; PRT; 462 AA.
 AC Q8BCU0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE UL74 protein.
 OS Human cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SW490;
 RX MEDLINE=22255414; PubMed=12368327;
 RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
 RT "The genes encoding the gCIII complex of human cytomegalovirus exist
 in highly diverse combinations in clinical isolates.";
 RL J. Virol. 76:10841-10848(2002).
 DR EMBL; AF531349; AAN40073.1; -;
 SQ SEQUENCE 462 AA; 53738 MW; 8FAAAA0416B33D06 CRC64;
 Query Match 83.7%; Score 2050.5; DB 2; Length 462;
 Best Local Similarity 85.0%; Pred. No. 7.7e-150;
 Matches 397; Conservative 23; Mismatches 40; Indels 7; Gaps 5;
 Qy 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWGLVLSKIGKYKLDQKL 60
 Db 1 MGRKGM-RGVNLFPLMSLTFLFSPINCAAVRLSVGRVSGKVLSTIGKQKLDKFKL 59
 Qy 61 EILRQLETTISTKY-NVSKQPVKNLTWN-TEFPQYIILAGPIQNYSTIYLWFDYFYSTQLR 118

Db 60 EILKQLEKDIYKYFNTRQHINKLWNMTFFPRYIILAGPIQNNSVTYLWFDYFSTQLR 119
Qy 119 KPAKYVYSQVNHAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTNPMF 178
Db 120 KPAKYVSEYNHTAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGN-TTNPMF 178
Qy 179 FNVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLVNAMSRLPRVPK 238
Db 179 FNVPRWNTKLYVGSKKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLVNAMSRLFRVPK 238
Qy 239 YINGTKLNTMRKLKRQAPVKEQFEKKAKKTQSTTTYPFYSYTTSAALNVTNTVYSITT 298
Db 239 YINGTKLNTMRKLKRQAPVKEQSEKKSKSQSTTTYPSPYTTTALNVTNTATYSVTT 298
Qy 299 AARVSTSTIAYRPDSSPFMSIMATQRLDLATWYTTILRYRONPFCPSRNRNAVSEPMK 358
Db 299 TARRISTSTIAYRPDSSPFMSIMTTQRLDLATWYTTILRYRONPFCSSRNRNAVSEPMK 358
Qy 359 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 359 NTHVLIRNETPYTYIGTLDMSLYNETMFVENKTASE--TPTSPSTGFQRTFIDPLW 415
Qy 419 DYLDLSLFLDEIRNFSLRSPYVNLTPPEHRRVNLSTNSLWMLQ 465
Db 416 DYLDLSLFLDEIRNFSLSQPTYGNLTPEHRRVNLSTNSLWMLQ 462

Search completed: December 7, 2004, 15:20:35
Job time : 202 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:12:05 ; Search time 39 Seconds
(without alignments)
790.715 Million cell updates/sec

Title: US-09-942-146A-1
Perfect score: 2451
Sequence: 1 MGRKEMVMDVPKMFVLISL.....PEHRAVNLSTNSLWHLQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2419	98.7	465	4	US-09-627-986-1
2	107.5	4.4	1168	1	US-08-620-117A-9
3	106.5	4.3	1167	1	US-08-485-568A-6
4	106.5	4.3	1167	2	US-08-590-554A-6
5	106.5	4.3	1167	2	US-09-184-223-6
6	105	4.3	452	4	US-09-543-681A-7874
7	101.5	4.1	351	4	US-09-248-796A-23675
8	101	4.1	476	4	US-09-248-796A-18994
9	99.5	4.1	978	2	US-08-415-593-43
10	97.5	4.0	1174	4	US-09-538-092-447
11	96.5	3.9	347	3	US-09-097-889-14
12	96.5	3.9	347	4	US-09-098-079-14
13	95.5	3.9	651	4	US-09-107-532A-4902
14	95.5	3.9	651	4	US-09-134-000C-5699
15	95	3.9	1065	4	US-09-538-092-84
16	95	3.9	2870	4	US-09-479-467A-15
17	95	3.9	3178	4	US-09-479-467A-4
18	94.5	3.9	354	4	US-09-248-796A-15271
19	94.5	3.9	383	4	US-09-248-796A-16746
20	94.5	3.9	515	4	US-09-248-796A-26713
21	94.5	3.9	705	4	US-09-248-796A-19971
22	94.5	3.9	844	4	US-09-991-258-19
23	94.5	3.9	2938	5	PCT-US94-00198-3
24	94	3.8	441	4	US-09-248-796A-23208
25	94	3.8	576	4	US-09-248-796A-20509
26	94	3.8	629	4	US-09-107-532A-6656
27	93.5	3.8	213	4	US-09-538-092-1255

28	93	3.8	541	4	US-09-248-796A-18318	Sequence 18318, A
29	92.5	3.8	537	4	US-09-248-796A-24443	Sequence 24443, A
30	92.5	3.8	1579	3	US-08-755-587-184	Sequence 184, App
31	92	3.8	288	4	US-09-216-393B-341	Sequence 341, App
32	92	3.8	288	4	US-09-216-393B-344	Sequence 344, App
33	91.5	3.7	250	4	US-09-248-796A-15562	Sequence 15562, A
34	91.5	3.7	354	3	US-09-307-925-8	Sequence 8, Appli
35	91.5	3.7	354	4	US-09-991-582B-8	Sequence 8, Appli
36	91.5	3.7	544	3	US-08-559-397A-30	Sequence 30, Appli
37	91	3.7	365	4	US-09-248-796A-15803	Sequence 15803, A
38	91	3.7	2186	2	US-08-822-445-2	Sequence 2, Appli
39	91	3.7	2186	3	US-09-396-540-2	Sequence 2, Appli
40	90.5	3.7	360	3	US-08-875-811-47	Sequence 47, Appli
41	90.5	3.7	721	4	US-09-248-796A-18800	Sequence 18800, A
42	90.5	3.7	797	4	US-09-134-000C-4997	Sequence 4997, Ap
43	90.5	3.7	1533	1	US-08-623-679-9	Sequence 9, Appli
44	90.5	3.7	1533	3	US-08-933-774-9	Sequence 9, Appli
45	90.5	3.7	1533	3	US-09-181-030-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1

US-09-627-986-1

; Sequence 1, Application US/09627986

; Patent No. 6569616

; GENERAL INFORMATION:

; APPLICANT: Compton, Teresa

; APPLICANT: Huber, Mary T.

; TITLE OF INVENTION: HUMAN CYTOMEGALOVIRUS GLYCOPROTEIN O AS A NEW DRUG

; FILE REFERENCE: 960296.96579

; CURRENT APPLICATION NUMBER: US/09/627.986

; CURRENT FILING DATE: 2000-07-28

; PRIOR APPLICATION NUMBER: 60/146,180

; PRIOR FILING DATE: 1999-07-29

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 1

; LENGTH: 465

; TYPE: PRT

; ORGANISM: Human cytomegalovirus

; US-09-627-986-1

Query Match	98.7%	Score 2419;	DB 4;	Length 465;
Best Local Similarity	99.4%	Pred. No. 7.6e-235;		
Matches 463;	Conservative 0;	Mismatches 1;	Indels 2;	Gaps 2;
Qy	1	MGRKEMVMDVPKMFVLISLISFLVSPINCKWMSKALYNRPWRGLVLSKIGKYKLDQLKL	60	
Db	1	MGRKEMVMDVPKMFVLISLISFLVSPINCKWMSKALYNRPWRGLVLSKIGKYKLDQLKL	60	
Qy	61	EILRQLETTISTKYNYSKQPKVNLTMNTEPPQYILAGPIONYSITVLPDPYSTQLRKP	120	
Db	61	EILRQLETTISTKYNYSKQPKVNLTMNTEPPQYILAGPIONYSITVLPDPYSTQLRKP	120	
Qy	121	AKVYSQYNHTAKTITPR-PPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF	179	
Db	121	AKVYSQYNHTAKTITPRPPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF	180	
Qy	180	NVPRWNTKLVGPTKVNVDSTQTYFLGLTALLRYAQNRCTHSPYLVNMSRNLFVRPKY	239	
Db	181	NVPRWNTKLVGPTKVNVDSTQTYFLGLTALLRYAQNRCTHSPYLVNMSRNLFVRPKY	240	
Qy	240	INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTTTFYSYTTSAALNVTNTVYSITTA	299	
Db	241	INGTKLKNTRKLRKQAPVKEQPEKAKKTQSTTTTTFYSYTTSAALNVTNTVYSITTA	300	
Qy	300	ARRVSTSTIAYRPDSSPFMSIMATQRLDALTAVVTLRYQNPFCEPSRNTAVSEFMKN	359	
Db	301	ARRVSTSTIAYRPDSSPFMSIMATQRLDALTAVVTLRYQNPFCEPSRNTAVSEFMKN	359	

QY 360 THVLIRNETPTIYGTLDSSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLMD 419
Db 360 THVLIRNETPTIYGTLDSSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLMD 419
QY 420 YLDSLLFLDEIRNPSLRSPYVNLTPPEHRAVNLSTNSLWMLQ 465
Db 420 YLDSLLFLDEIRNPSLRSPYVNLTPPEHRAVNLSTNSLWMLQ 465
RESULT 2
US-08-620-717A-9
; Sequence 9, Application US/08620717A
; Patent No. 5670365
; GENERAL INFORMATION:
; APPLICANT: Feitelson, Jerald S.
; TITLE OF INVENTION: Identification of, and Uses For, Nematocidal
; TITLE OF INVENTION: Bacillus thuringiensis Genes, Toxins, and Isolates
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/620,717A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/540,104
; FILING DATE: 06-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA94.C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1168 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: 167p
US-08-620-717A-9
Query Match 4.4%; Score 107.5; DB 1; Length 1168;
Best Local Similarity 20.9%; Pred. No. 0.22;
Matches 97; Conservative 68; Mismatches 131; Indels 169; Gaps 26;
QY 90 FPOY-ILAGPQNYSTYLVWFDYSTQLRKPAKYVSYQYNNHTAKTITRPPPCGRVPSM 148
Db 8 YSYHNVLHPIFLDS----FDDPP-VETFKDLKGAWBEF-----GKTGYM 48
QY 149 TCSLEMLNYSKRNDSGEQCGNFTTNPMFFNPRWTKLYVGPVKVND-----SQ 200
Db 49 DPLKQHLQIAW--DTSQNG-----TVDYIALTKASISLGLIPGADA 88
QY 201 TIYFELGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYNG- 242
Db 89 VPFPIINFDVIFPKLFGSGSQQAQAQPFELIIEKVELDVEDFENFTLNLLNLDGM 148
QY 243 ----TKLKNTR-----KLKRKOAP-----VKEQPEKAKKTQSTTTPY 277

Db 149 QVALSHFQNDVQIAICQGEQPLMDQTPACTTTHLISVRESF-KDARTTIETALPH 207
QY 278 FS-----YTTSAALNVTNTVYITTAARVSTSTIAYRPD 313
Db 208 FKXPMLSNDNDPDPNSDVLTLTPMYTTAATNLILHQGY-IQFAERWKSVMY-----D 261
QY 314 SSFM--KSIMATQLRDLATWYTTLYRQNPFCPPSRNRTAVSBFMKXNTHVLIINET-- 368
Db 262 ESFINQTKVDLQRIQDYSTTVSTTTFE-KFKPTLNFN-NKESVKNYR-----YVRSMTLQ 315
QY 369 -----PYIYGTLDMSLYNETMFV-----ENKTASDNKTTPTSPSMG 408
Db 316 SLDIAATWPTLDNVNVPNSVDIQLDQTRLVFSDVAPMEGNDNIYSNIIDVLTPINTGIG 375
QY 409 FQ-----RTFIDPLMDYLDLSLLFLDEIRNFSRLSPYVNLTPPEH 448
Db 376 FOESSDLRKFTTPRIE-LQSMQFHGQ-----YVNSKSVEH 409
RESULT 3
US-08-485-568A-6
; Sequence 6, Application US/08485568A
; Patent No. 5589382
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fu, Jenny M.
; TITLE OF INVENTION: No. 5589382el Bacillus thuringiensis Genes Encoding
; TITLE OF INVENTION: Nematode-Active Toxins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,568A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/310,197
; FILING DATE: 21-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/918,345
; FILING DATE: 21-JUL-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/558,738
; FILING DATE: 27-JUL-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,698
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,403
; FILING DATE: 30-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/999,053
; FILING DATE: 14-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.

```
;
;
;   REGISTRATION NUMBER: 31,794
;   REFERENCE/DOCKET NUMBER: MA48DD2.C1
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (904) 375-8100
;   TELEFAX: (904) 372-5800
;   INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1167 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-485-568A-6

Query Match          4.3%; Score 106.5; DB 1; Length 1167;
Best Local Similarity 20.9%; Pred. No. 0.28;
Matches 97; Conservative 67; Mismatches 132; Indels 169; Gaps 26;

Qy 90 PPQY-ILAGPIQNSITYLWDFYSTQLRKPQKVVYQVNHAKTITFRPPCGRVPSM 148
Db 8 YPSYHNVLAHPRLDS-----FFDPF-VETFKDLGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTNPMPFNVPWMTKLVGPTKVNVD-----SQ 200
Db 49 DPLKQHLQIAW--DTSNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYPLGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYING- 242
Db 89 VVPFINMFVDFIPKLFGRGSGQNAQAQFFELIIEKVELVDEDFRFTLNLLNLYDGM 148
Qy 243 -----TKLNTWR-----KLKQKQAP-----VKEQPEKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQPGLMDQTPTACTPTDHLISVRESF-KDARTTIETALPH 207
Qy 278 FS-----YTTSAALNVTTNVTYSITTAARRVSTSTIAVRPD 313
Db 208 FKNPMLSTNDNTPDFNSDTVLLTLPMYTTGATLNLILHQGY-IQFAERWKSVMY-----D 261
Qy 314 SSFW---KSMATQLRLATVYTTLYRQNPFCPSRRNRTAVSEFMKNTHLVIRNET-- 368
Db 262 ESFINQTKVLDLORRIQDYSITVSTTFE-KFKPTLNPS-NKESVVKYNR-----YVRSMTLQ 315
Qy 369 -----PYTIYGLDMSLSLYNMTFV-----ENKVASDSNKTTPSPSMG 408
Db 316 SLDIAATWPTLDNVNPSNVDIQLDQTRLVPSDVAGPWEGNDNITSNIIDVLTPTINGIG 375
Qy 409 FQ-----RTIDPLWDYLDLSLLFLDEIRNFSRLSPTVYVNLTPPEH 448
Db 376 FOESSDLRKFTYPRIE-LQSMQPHGQ-----YVNSKSVSH 409
```

```
RESULT 4
US-08-590-554A-6
; Sequence 6, Application US/08590554A
; Patent No. 5831011
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel
; APPLICANT: Narva, Kenneth E.
; APPLICANT: Fu, Jenny M.
; TITLE OF INVENTION: No. 5831011el Bacillus thuringiensis Genes Encoding
; TITLE OF INVENTION: Nematode-Active Toxins.
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
;
;
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/590,554A
;   FILING DATE: 21-MAR-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/485,568
;   FILING DATE: 7-JUN-1995
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/310,197
;   FILING DATE: 21-SEP-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/918,345
;   FILING DATE: 21-JUL-1992
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/558,738
;   FILING DATE: 27-JUL-1990
;   CLASSIFICATION: 435
;   APPLICATION NUMBER: US 08/357,698
;   FILING DATE: 16-DEC-1994
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/176,403
;   FILING DATE: 30-DEC-1993
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/999,053
;   FILING DATE: 14-JUL-1993
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Saliwanchik, David R.
;   REGISTRATION NUMBER: 31,794
;   REFERENCE/DOCKET NUMBER: MA48DD2.C2
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (352) 375-8100
;   TELEFAX: (352) 372-5800
;   INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1167 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;
US-08-590-554A-6

Query Match          4.3%; Score 106.5; DB 2; Length 1167;
Best Local Similarity 20.9%; Pred. No. 0.28;
Matches 97; Conservative 67; Mismatches 132; Indels 169; Gaps 26;
```

```
Qy 90 PPQY-ILAGPIQNSITYLWDFYSTQLRKPQKVVYQVNHAKTITFRPPCGRVPSM 148
Db 8 YPSYHNVLAHPRLDS-----FFDPF-VETFKDLGAWEEF-----GKTGYM 48
Qy 149 TCLSEMLNVSKRNDTGEQCGNFTTNPMPFNVPWMTKLVGPTKVNVD-----SQ 200
Db 49 DPLKQHLQIAW--DTSNG-----TVDYLAUTKASISLIGLIPGADA 88
Qy 201 TIYPLGL-----TALLRYAQRNCTHSFY-----LVNMSRN--LFRVPKYING- 242
Db 89 VVPFINMFVDFIPKLFGRGSGQNAQAQFFELIIEKVELVDEDFRFTLNLLNLYDGM 148
Qy 243 -----TKLNTWR-----KLKQKQAP-----VKEQPEKAKKTOSTTTPY 277
Db 149 QTALSHFQNDVQIAICQEQPGLMDQTPTACTPTDHLISVRESF-KDARTTIETALPH 207
Qy 278 FS-----YTTSAALNVTTNVTYSITTAARRVSTSTIAVRPD 313
Db 208 FKNPMLSTNDNTPDFNSDTVLLTLPMYTTGATLNLILHQGY-IQFAERWKSVMY-----D 261
```


ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/097,889
FILING DATE: 15-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman Ph.D., Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 660088.417
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 347 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-097-889-14

Query Match 3.9%; Score 96.5; DB 3; Length 347;
Best Local Similarity 22.3%; Pred. No. 0.43;
Matches 68; Conservative 33; Mismatches 109; Indels 95; Gaps 12;

QY 170 NFFTPNMFNVPRWNTKLYVGPVKVNDVSTIYFLGJTALL-----RYAQRNCTH-- 221
DB 36 NMLAFIPVL--TKQNPSTEAALKYFTQATASHMILLMAILFNMLSGQWMTNTTQY 93

QY 222 -SFVLVNAISRNL-----FRVPKYINGTKLNTMRKLK-RKQAPVKEQFEKAKKTOST 273
DB 94 SSLMIMMAMKMGWAPFHWPEVTQGTPLTSLGLLLTWQKLAPIS----- 140

QY 274 TTPFSYTTSAALNVTNTVYSITTAARRVSTSTIAYRPSDSFMSIMA-----TOL 325
DB 141 ----IMYQISPLSNVSLLLTSLIL-----SIMAGSWGGLNQTL 175

QY 326 RDLAT-----WVYTTLYRQNPFCPSRNRRTAVSEFMKNTVLIENETPYTIYGLTD 377
DB 176 RKILAYSSITHGMMAVLPY-----NPNNTILNLTIIYILTTTA 215

QY 378 MSSLYYNE--TMFVENKTASDNKTTPTSPSMGFORTFIDPLWDYDLSLLFLDEI-RNFS 434
DB 216 FLLNLNSSTTLLLSRTWKNLTLPLIPSTLLSLGGLPLTGLPKWAIIEFTKNS 275

QY 435 LRSPT 439
DB 276 LIPT 280

RESULT 12
US-09-098-079-14
Sequence 14, Application US/09098079
Patent No. 6489095
GENERAL INFORMATION:
APPLICANT: Hernstadt, Corrina
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Clevenger, William
APPLICANT: Fahy, Eoin F.
APPLICANT: Davis, Robert E.
TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:

ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,079
FILING DATE: 15-JUN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rosenman Ph.D., Stephen J.
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 660088.416
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 347 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-098-079-14

Query Match 3.9%; Score 96.5; DB 4; Length 347;
Best Local Similarity 22.3%; Pred. No. 0.43;
Matches 68; Conservative 33; Mismatches 109; Indels 95; Gaps 12;

QY 170 NFFTPNMFNVPRWNTKLYVGPVKVNDVSTIYFLGJTALL-----RYAQRNCTH-- 221
DB 36 NMLAFIPVL--TKQNPSTEAALKYFTQATASHMILLMAILFNMLSGQWMTNTTQY 93

QY 222 -SFVLVNAISRNL-----FRVPKYINGTKLNTMRKLK-RKQAPVKEQFEKAKKTOST 273
DB 94 SSLMIMMAMKMGWAPFHWPEVTQGTPLTSLGLLLTWQKLAPIS----- 140

QY 274 TTPFSYTTSAALNVTNTVYSITTAARRVSTSTIAYRPSDSFMSIMA-----TOL 325
DB 141 ----IMYQISPLSNVSLLLTSLIL-----SIMAGSWGGLNQTL 175

QY 326 RDLAT-----WVYTTLYRQNPFCPSRNRRTAVSEFMKNTVLIENETPYTIYGLTD 377
DB 176 RKILAYSSITHGMMAVLPY-----NPNNTILNLTIIYILTTTA 215

QY 378 MSSLYYNE--TMFVENKTASDNKTTPTSPSMGFORTFIDPLWDYDLSLLFLDEI-RNFS 434
DB 216 FLLNLNSSTTLLLSRTWKNLTLPLIPSTLLSLGGLPLTGLPKWAIIEFTKNS 275

QY 435 LRSPT 439
DB 276 LIPT 280

RESULT 13
US-09-107-532A-4902
Sequence 4902, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts

Search completed: December 7, 2004, 15:22:19
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:21:31 ; Search time 164 Seconds
(without alignments)
1017.129 Million cell updates/sec

Title: US-09-942-146A-1

Perfect score: 2451

Sequence: 1 MGRKEMVRDVPFMFVLISI.....PHRRAVNLTSGNSLWMLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Gap extension penalty reduced to zero to allow for deletion.

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_23Sep04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2451	100.0	465	ADO15374	CMV glyco
2	2420	98.7	465	ADBB1330	Human cyt
3	710	29.0	31267	ABG74786	Human RGS
4	681	27.8	26926	ADQ17316	Human sof
5	678	27.7	26926	AAU05396	Human tit
6	581	23.7	16368	ABM67171	Phototrab
7	580	23.7	11300	ABB84277	Adenoviru
8	575	23.5	15281	AAR44929	T. niveum
9	561	22.9	22157	ADP84155	Human CAL
10	539	22.0	8601	AAO29660	Paederus
11	538	22.0	19938	ABP76678	Streptomy
12	536	21.9	19938	ABP76681	Streptomy
13	534	21.8	19938	ABP76680	Streptomy
14	531	21.7	19938	ABP76679	Streptomy
15	526	21.5	7107	ABBS58144	Drosophil
16	522	21.3	19938	ABBS98398	Streptomy
17	518	21.1	6940	ADN96828	Bugula br
18	516	21.1	10182	ADP38334	Staphyloc
19	514	21.0	10203	ADO84851	S. epiderm
20	508	20.7	19938	ABP76682	Streptomy
21	507	20.7	7052	AAE35266	Human P45
22	499	20.4	11721	ABUS4721	Human CAL
23	496	20.2	7718	ABG11811	Novel hum
24	492	20.1	6619	ABG23329	Novel hum
25	490	20.0	7036	AAE35264	Human P45

26	489	20.0	7067	6	AAE35265	Human P45
27	488	19.9	7368	6	ABG76187	Human ser
28	483	19.7	9222	4	ABG21064	Novel hum
29	483	19.7	11877	8	ADL91934	Streptomy
30	483	19.7	12199	3	AAV77180	S. venezu
31	480	19.6	5909	4	ABG23295	Novel hum
32	476	19.4	9799	6	ABUS4708	Human CAL
33	475	19.4	10421	6	ABU33707	Protein e
34	473	19.3	6266	6	AAO29662	Paederus
35	473	19.3	8805	4	ABB67112	Drosophil
36	472	19.3	10917	6	AAE36132	Streptomy
37	469	19.1	5546	5	AAU85008	Human mel
38	469	19.1	10431	6	ABUS4861	Human CAL
39	468	19.1	5746	5	AAU84596	HIV compl
40	468	19.1	6025	4	AAG84939	Shrimp wh
41	466	19.0	7064	7	ADG70546	Aspergill
42	465	19.0	4688	6	ABU48941	Protein e
43	465	19.0	5005	6	ABU48947	Protein e
44	465	19.0	5266	4	ABG08561	Novel hum
45	465	19.0	10498	6	ABJ19119	Pathogen

ALIGNMENTS

RESULT 1

ADO15374

ID ADO15374 standard; protein; 465 AA.

XX

AC ADO15374;

XX

DT 01-JUL-2004 (first entry)

XX

DE CMV glycoprotein O (go) polypeptide.

XX

KW CMV; glycoprotein O; go; glycoprotein H ; glycoprotein L; glycoprotein B;

KW CMV infection; tumour; cancer; virucide.

XX

OS Cytomegalovirus.

XX

PN US2004013682-A1.

XX

PD 22-JAN-2004.

XX

PF 26-JUN-2002; 2002US-00942146.

XX

PR 29-JUL-1999; 99US-0146180P.

PR

28-JUL-2000; 2000US-00627986.

XX

PA (COMP/) COMPTON T.

PA

(HUBER/) HUBER M T.

XX

PI Compton T, Huber MT;

XX

DR WPI; 2004-121532/12.

XX

Designing new anti-CMV drug by analyzing binding of glycoprotein O to

glycoprotein O receptor, designing candidate drug that would

competitively interfere with glycoprotein O binding to glycoprotein O

receptor.

XX

PS Abp76679 Streptomy

XX

PS Abp76680 Streptomy

XX

PS Abbs58144 Drosophil

XX

CC Abbs98398 Streptomy

CC

CC Adn96828 Bugula br

CC

CC Abp38334 Staphyloc

CC

CC Ado84851 S epiderm

CC

CC Abp76682 Streptomy

CC

CC Aae35266 Human P45

CC

CC Abus4721 Human CAL

CC

CC Abg11811 Novel hum

CC

CC Abg23329 Novel hum

CC

CC Aae35264 Human P45

CC

Disclosure; Fig 3; 7pp; English.
The invention relates to a method of designing a new anti-cytomegalovirus (CMV) drug, involving analysing the binding of glycoprotein O to a glycoprotein O receptor, designing a candidate drug that would competitively interfere with glycoprotein O binding to the glycoprotein O receptor and showing that the candidate drug competitively inhibits glycoprotein O binding to the glycoprotein O receptor. The invention also relates to a method of screening involving determining whether a candidate drug interferes with a glycoprotein O-containing complex binding to a cell surface. The evaluation involves analysing the inhibition of major early protein of the human cytomegalovirus (HCMV).

CC The invention also relates to a vaccine useful for diminishing CMV
 CC infection and comprising at least a fragment of the glycoprotein O
 CC polypeptide in combination with a carrier. The vaccine comprises at least
 CC a fragment of a glycoprotein chosen from CMV glycoprotein H, CMV
 CC glycoprotein L, or CMV glycoprotein B, and is useful for vaccinating a
 CC patient against CMV infection. The drug and the vaccine are useful for
 CC treating CMV infections in tumours. This sequence represents the CMV
 CC glycoprotein O (gO) polypeptide of the invention.

XX Sequence 465 AA;

Query Match 100.0%; Score 2451; DB 8; Length 465;
 Best Local Similarity 100.0%; Pred. No. 2.6e-63;
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGRKEMVRDVPKMFVLISIFLVSFNCVKMSKALYNRPWRGLVLSKIGYKLDQL 60
 DB 1 MGRKEMVRDVPKMFVLISIFLVSFNCVKMSKALYNRPWRGLVLSKIGYKLDQL 60
 QY 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEFPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120
 DB 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEFPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120
 QY 121 AKYVYSQNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFFN 180
 DB 121 AKYVYSQNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFFN 180
 QY 181 VPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEMLNVSKENDTGEQCGNFTTFNPMFFN 240
 DB 181 VPRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEMLNVSKENDTGEQCGNFTTFNPMFFN 240
 QY 241 NGTKLNTMKRLKQAPVKEQEKAKKTQSTTTPYFSTTSAALNVTNTVYSITTA 300
 DB 241 NGTKLNTMKRLKQAPVKEQEKAKKTQSTTTPYFSTTSAALNVTNTVYSITTA 300
 QY 301 RRVSTSTIAYRPDSSFMKSIMATQRLDLATVVTTLRYQRNPFCEPSRNRATVSEFMKN 360
 DB 301 RRVSTSTIAYRPDSSFMKSIMATQRLDLATVVTTLRYQRNPFCEPSRNRATVSEFMKN 360
 QY 361 HVLRNETPTVYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWDY 420
 DB 361 HVLRNETPTVYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWDY 420
 QY 421 LDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465
 DB 421 LDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465

RESULT 2
 ID ADB81330

XX ADB81330 standard; protein; 465 AA.

AC ADB81330;

DT 04-DEC-2003 (first entry)

XX Human cytomegalovirus glycoprotein O polypeptide.

XX glycoprotein O; gO; anti-cytomegalovirus; CMV; AIDS;
 XX bone marrow transplant; glycoprotein H; glycoprotein L; gene therapy.

OS Human herpesvirus 5.

XX US6569616-B1.

PN 27-MAY-2003.

PD 28-JUL-2000; 2000US-00627986.

PF 29-JUL-1999; 99US-0146180P.

PR (WISC) WISCONSIN ALUMNI RES FOUND.

XX

PI Compton T, Huber MT;

XX WPI; 2003-605462/57.

XX Screening candidate anti-cytomegalovirus (CMV) drugs for the ability to
 PT block CMV entry into a host cell comprises determining whether a
 PT candidate drug interferes with glycoprotein O binding to glycoprotein O
 PT receptor.

XX Disclosure; Fig 3; 8pp; English.

XX This invention relates to a novel method of designing a new anti-
 CC cytomegalovirus (CMV) drug that has the ability to block CMV entry into a
 CC host cell. CMV is a member of the Herpesviridae family and is an
 CC opportunistic pathogen responsible for serious clinical disorders in
 CC immunosuppressed patients, such as persons with AIDS or those with recent
 CC organ or bone marrow transplants. The present invention describes a new
 CC drug target that comprises a CMV viral component designated glycoprotein
 CC O (gO), whereby the candidate drug works by interfering with gO binding
 CC to its receptor, where glycoprotein O is part of a complex with
 CC glycoprotein H and glycoprotein L. Accordingly, using gene therapy or
 CC developing antibodies to go that act as vaccines, it is possible to block
 CC virus infection by impeding the viral entry pathway. Furthermore, the
 CC method is useful in identifying alternative drug targets and immunogens
 CC that elicit protective activity against CMV infection. This polypeptide
 CC is the human cytomegalovirus glycoprotein O amino acid sequence of the
 CC invention.

XX Sequence 465 AA;

Query Match 98.7%; Score 2420; DB 7; Length 465;
 Best Local Similarity 99.4%; Pred. No. 2.2e-62;
 Matches 463; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
 QY 1 MGRKEMVRDVPKMFVLISIFLVSFNCVKMSKALYNRPWRGLVLSKIGYKLDQL 60
 DB 1 MGRKEMVRDVPKMFVLISIFLVSFNCVKMSKALYNRPWRGLVLSKIGYKLDQL 60
 QY 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEFPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120
 DB 61 EILRQLETTISTKYNVSKQPVKNLTMTNTEFPQYVILAGPIQNYSTIYLFDFYSTQLRKP 120
 QY 121 AKYVYSQNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFF 179
 DB 121 AKYVYSQNHHTAKTITPRPPCGRVPSMTCLSEMLNVSKENDTGEQCGNFTTFNPMFF 180
 QY 180 NVRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEMLNVSKENDTGEQCGNFTTFNPMFF 239
 DB 180 NVRWNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSEMLNVSKENDTGEQCGNFTTFNPMFF 240
 QY 240 INGTLKNTMKRLKQAPVKEQEKAKKTQSTTTPYFSTTSAALNVTNTVYSITTA 299
 DB 240 INGTLKNTMKRLKQAPVKEQEKAKKTQSTTTPYFSTTSAALNVTNTVYSITTA 300
 QY 300 ARRVSTSTIAYRPDSSFMKSIMATQRLDLATVVTTLRYQRNPFCEPSRNRATVSEFMKN 359
 DB 300 ARRVSTSTIAYRPDSSFMKSIMATQRLDLATVVTTLRYQRNPFCEPSRNRATVSEFMKN 359
 QY 360 THVLIRNETPTVYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 419
 DB 360 THVLIRNETPTVYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGQRTFIDPLWD 419
 QY 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465
 DB 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRRVAVNLSTNSLWMLQ 465

RESULT 3
 ADB74786

ID ADB74786 standard; protein; 31267 AA.

XX

AC ADB74786;

XX

DT	05-JUN-2003	(first entry)	12	-----PK-----	13
XX	DE	Human RGS11 protein.			
XX	KW	RGS11; human; screening; cardiant; antiangiinal; gene therapy;	629	SLTEQAGPEPAAPYFITKPVVQKLVEGGSVVFGQVGNPKPRHYWKSGVPLTTGYR	688
XX	KW	heart disorder; cardiac ischaemia; heart failure; angina.	14	-----	13
XX	KW		689	YKVSYNKQTGECKLVISMTFADDAEYTIIVRNKHGETSASASLLEADYELLKMSQEM	748
OS	XX	Homo sapiens.	14	-----	13
XX	PN	WO2002103355-A1.	749	LYQTQVTAFOBPEVGETAPGFVYSEYEKEQALIRKMAKDTVVVRYVEDQEFHI	808
XX	PD	27-DEC-2002.	14	-----	13
XX	XX	17-JUN-2002; 2002WO-JP006019.	809	SSFEERLIKEIYRIIKTTLEELLEDGEBKMAVDISESEAVESGFDLRKRYRILEGMG	868
XX	XX	18-JUN-2001; 2001JP-00183038.	14	-----	13
XX	PA	(TAKE) TAKEDA CHEM IND LTD.	869	CHFSLQDVWISITKDOKRIKHGERYQMDFLQDGRASLRIPVVLPEDEGIYTAFASNIKGN	928
XX	PI	Koyama N, Tanida S, Yamamoto K;	14	-----	13
XX	XX	WPI; 2003-167557/16.	929	AICSGKLYVEPAAPLGAFTYIPTLEBPVSRIISLSPRSVSRSPIRMSPARMSPARMSPARM	988
DR	N-PSDB; ABX13540.		14	-----MFVLISISF-----	22
XX	XX	Screening compounds regulating RGS11 expression and activity for	989	SPARMSPGRRLEBETDSQLERLYKPVFLKPVSKFCLEGQTARFDLKVGRPMPTFWPH	1048
PT	prevention and treatment of heart disease.		23	-----	22
XX	PS	Claim 1; Page 59-261; 321pp; Japanese.	1049	DGQOIWVNDYTHKVIKEDGTQSLIIVPATPSDGEVTVVQNRAGRSSISVILTVEAVEH	1108
CC	CC	This invention describes a novel method for screening compounds for their	23	-----	22
CC	CC	ability to regulate the activity and expression of human RGS11 and its	1109	QVKPMPEKLVKNVNIKEGSRLEMKVATGNPNPDIVLKNSDIIVPHKYPKIRIEGTKE	1168
CC	CC	partial peptides and salts, by observing the expression or activity of	23	-----LLV-----	25
CC	CC	RGS11 in the presence or absence of the test compound. The products of	1169	AALKIDSTVSQDSAWYTATATAINKAGRDTTTRCKVNVVEFAEPERPERKLIIPRGTYRAKEI	1228
CC	CC	the invention have cardiant and antiangiinal activity and can be used for	26	-----	25
CC	CC	gene therapy. The methods and compositions are useful in the prevention,	1229	AAPELEPLHLRYGQEQWEGDLYDKEKQKPFKKLTSLRLKRFPAHPECLTPIGDP	1288
CC	CC	treatment and diagnosis of heart disorders such as cardiac ischaemia,	26	-----SFINCKVMSK-----	35
CC	CC	heart failure and angina. This sequence represents the human RGS11	1289	TWVVEHLHDGKPLEAANRLRMINEFGYCSLDYGVAYSRSDSGIITCRATNKYGTDHTSATL	1348
CC	CC	protein described in the disclosure of the invention	36	-----ALYNRPWR- 43	
XX	SQ	Sequence 31267 AA;	1349	IVKDEKSLVEESQLPEGRKGLORIEBELERMAHEGALTGVTTDOKEKQKPDIVLYPEPVRV	1408
			44	-----	43
			1409	LEGETARFCRVGTGYPQPKVNMVNLQGLIRKSRFRVRYDGIHLDVCKSYDTGEVKV	1468
			44	-----GLVLSKI-----	50
			1469	TAENPEGVIEHKVLEIQOQREDFRSVLRAPRPRPFHVHEPGKLFQEVQKVDPRVDTE	1528
			51	-----	50
			1529	TKEVVKLKRABRITHEKVPSESEELRSKFRRTTEGYEAITAVELKSRKDESEBELR	1588
			51	-----	50
			1589	KTKDELLHWTKLTBEEKKALAEKGKITTTPKPKDIELSPSMEAPKIFERIQSQTGQG	1648
			51	-----	50
			1649	SDAHFRVVRVKGPDPECEWYKNGVKIERSDRIYVWYWPEDVNCVCELVIRDVTAEDSASIMVK	1708
			51	-----	50

Query Match 29.0%; Score 710; DB 6; Length 31267;
 Best Local Similarity 0.9%; Pred. No. 2.5e-07;
 Matches 271; Conservative 92; Mismatches 98; Indels 30346; Gaps 87;

QY	3	RKE-----	5
DB	209	RKEAKTALSTIAVATAKAKEQETILTRETMTATROEQIQTHTGKVDVGGKAEAVTVVA	268
QY	6	-----	5
DB	269	AVDQARVREPREPHLEESYAQQTILEYGYKERISAQVAEPPQRPASBPHVVPKAVKPR	328
QY	6	-----	5
DB	329	VIQAPSETHIKTTDQGMHISSQIKKTTDLTTERLVHVDKRPRTASPHFTVSKISVPKTE	388
QY	6	-----	5
DB	389	HGYEASIGASAIATLOKELSATSQAOKIKYSVKAPTUKPSETVRABEPTLPQFPADTP	448
QY	6	-----	5
DB	449	DTYKSEAGVEVKEVGSITGTTVREERPEVLHGREAKVTETARVPAPVEIPTVPTPLVS	508
QY	6	-----MMVRDV-- 11	
DB	509	GLKNVTIEGESVLECHISGSPSTVTVWRYDYQIESIDFQITFQSGIARLMIREAFA	568
QY	12	-----	11
DB	569	EDSGRTCSAVNAGTVSTSCYLAQVSEEFKEKETTAVTEKFTTBKRFVESRDVMTDT	628

Db	1709	AINIAGETSSHAFLVQAKLITFTELOQDVVAKKOTMATFECETSEPFVKVWKYKDG	1768
Qy	51	-----	50
Db	1769	EVHGDKYRMHSDRKVHFLSILITDTSDAEDYSCVLVEDENVKTTAKLIVEGAVVEFVKE	1828
Qy	51	-----GKY-----KLD	56
Db	1829	LODIEVPESYSELECIVSPENIBGKWYHNDVELKSNQKTYITSRGRQNLTVKDVTKED	1888
Qy	57	Q-----	57
Db	1889	QGEYSFVIDGKTKTKLKMKPRPAILQGLSDQKVCBGDIVQLEVKVLSVESVEGVMKOG	1948
Qy	58	-----LKLILRL-----	66
Db	1949	QEVQPSDRVHIVIDKQSHMLLIJEDMTKEDAGNYSFTIPALGLSTSGRSVSVKIKILGLR	2008
Qy	67	-----ETTISTKYNVS-----	77
Db	2009	DLTCTQNVVFEVELSHGIDVLWNFKDKIEKPSKSKYIEAHGKIYKLTVLNMMKDEG	2068
Qy	78	-----	77
Db	2069	KYTFYAGENMTSGKLTVAGGAISKPLTDQTVASSQEAVFCEVANPDSKGWLRDGHLP	2128
Qy	78	-----KQPVKNLTM-----	86
Db	2129	LTNNIRTVKIKTKLNLTVTQDAVFTVELTHPNVKGQWIKNGVVLESNEKYAISVG	2188
Qy	87	-----	86
Db	2189	TIYSLRIKCAIVDESIVYFRLGLRGASARLHVETVKIIKKPKDVTALENATVAFEVSVS	2248
Qy	87	-----	86
Db	2249	HDTVPUKWFHKSVEIKPSDKHRLVSRKVKHKLMLQNISPSDAGEYTAUVGQLECKALFV	2308
Qy	87	-----NTEFPQ-----	92
Db	2309	ETLHITKMKNIIEVPEKTKASPECEVSHFNVPNSMWLKNQVEIEMSEKFKIVVQGLHQLI	2368
Qy	93	-----	92
Db	2369	IMNTSTEDSABYTFVCGNDQVSATLTVTPIMITSMCLKDINAEEKDTITFEVTVNYEGISY	2428
Qy	93	-----	92
Db	2429	KWLKNGVEIKSTDKQWRTKLTHSLIIRNVHFGDAADYTFVAGKATSTATLYVEARHIE	2488
Qy	93	-----	92
Db	2489	FRKHIDIKVLEKKRAMFECEVSEPDITVQMMKDDQELQITDRIKIOKEKYVHRLLIPT	2548
Qy	93	-----YVILAG-----	98
Db	2549	RMSDAGKYTVWAGNVSTAKLFVEGRDVRIRSIKKEVQVIEKQRAVVEFEVNEDDVDAMH	2608
Qy	99	-----	98
Db	2609	YKDGIEINFQVQRHKYVVERRIHRMFISETRQSDAGEYTFVAGNRSSVTLYVNAPEPP	2668
Qy	99	-----	98
Db	2669	QVLQELQVTVQSGKPARFCAMISGRPOPKISWYKEEQLLSTGFKCKFLHDGQEVTLILI	2728
Qy	99	-----PIQ-----	101
Db	2729	EAFPEDAAYTCEAKNDYGVATTSASLSVEFVSVSDQEMPVYPPIAITPLQDVTVTSEG	2788
Qy	102	-----	101

Db	2789	QPARFCRVSGTDLKWSYKSKDKKIKPSRFRMTQFEDTYQLEIAEAYPEDEGTYTFVAN	2848
Qy	102	-----	101
Db	2849	NAVQVQSSTANLSLEVSFKKEPLGQKPSFIQPLSSLRVNHGETVRFHARVSGIPKPEIQW	2908
Qy	102	-----	101
Db	2909	FHNQQLILPTKDVVHFPESTGMALMLIIVDAYSEHAGQYSCKAANSAGERTCAATLTVTP	2968
Qy	102	-----	101
Db	2969	KAASVQDTQLCHTASLSQIAESTELSKAKESTDDSFIDVTWTHGAKIEESERLKQSQ	3028
Qy	102	-----	101
Db	3029	NGNIQFLTICNVQLVDQGLYSICVHNDCGERTTSAVLSVEDVNVNTOARGOLEEIMANTT	3088
Qy	102	-----	101
Db	3089	IGESSTYITNVCEKEBVEYETPENVSQAITPHASESPGTLVNVVESEIEIASERIKKDDLRE	3148
Qy	102	-----	101
Db	3149	LQLSASTRIDEFKTEQKEENMRFENSRFRQRCPPSPLOEBIESQEVYEGDSCNPFCHF	3208
Qy	102	-----NYSITVIL-----	108
Db	3209	QGYQPQIVTWYNDMPPIPRNQNFIIHSLNYSILITSSVHHQNEGSITCVLFNQYGTVKT	3268
Qy	109	-----	108
Db	3269	TSMLKVKAKQKHDVKAHKVPVFDYLDDEEELAVDQAKGAHPSMSQEGQTNLHLLKTN	3328
Qy	109	-----	108
Db	3329	PPVPPSGDTLLSPFVBEIQVTAATPIPEQDKESKEVFQTELEPKAMPQDQVTPKGRF	3388
Qy	109	-----WF-----	110
Db	3389	VFLSDITNEPPKMLQEMPKHARCREGDSIILECLISGEPQPVVTVFQNGVLLKQNKQFQ	3448
Qy	111	-----	110
Db	3449	EEVNCSHQLYIKDVSQDSGKYKCAENNSGAVESDLTVEPVTYRENSQENIGEIVG	3508
Qy	111	-----	110
Db	3509	KYSRQQLQDQGESVRAHFYDYPAGPFTWTVNVKEYSVRDYFQSLEETIEQIDQKQREVP	3568
Qy	111	-----	110
Db	3569	ETKDRDQAKLVQSESIITSMEEVEVTFNTVVEYNNQKQESLGRPLSPESDISIGVSTTS	3628
Qy	111	-----DFYSTQL-----	117
Db	3629	BEISELDQYTPPSSVEYFESPKSPDLYFNPSDITKQSSIHSGGETVRYSTPLGEVAER	3688
Qy	118	-----	117
Db	3689	YSTPSEGEVGERYSTPPGGETLERYSTPPGGETLERYSTPPGGETLERYSTPPGGETLERYSTP	3748
Qy	118	-----	117
Db	3749	PGETLERYSTPPGALERYSIPTGPNPTGTFKTYPSKIEREDGTPNEHFFYPTTEERCSA	3808
Qy	118	-----	117
Db	3809	YEIWRSDSGFTPNBAIBPKONMPPSFIEPLTKRKVYENTTLGLFIVEGLPFGVKWYR	3868
Qy	118	-----	117
Db	3869	NKSILLEPDERIKMERVGNVCSLEISNIQKGGEGYCHAVNIIGEAKSFANVDIMPQER	3928

QY 118 -----RKPA----- 121
Db 3929 VVALPPVTHQWMEFDEHTTSRTSPQBI VLEVELSEKVKFEKQKVI VTVPEFTP 3988
QY 122 ----- 121
Db 3989 DHKSMIVSLDVLFPNFVDPNMDRGEDEKELKIDLEVFEMPPRFIMPICDFKIPENS DAV 4048
QY 122 -----KYVYSQ----- 127
Db 4049 FKCSVIGIPTPEVKWYKMYCIBPDNIKYVISEKGSHTLKIRNVCLSDSATYRCRAVNC 4108
QY 128 ----- 127
Db 4109 VGBAICRGFLTMGDSEIFAIAKSKVTLSSLMEELVKNSYTDSPFQVQVGGPPRFIK 4168
QY 128 ----- 127
Db 4169 GISDCVAPIEFSSPFLSABEGLHSABLQLSKINETLELSESPVYSTKFDSEKGTGPI 4228
QY 128 ----- 127
Db 4229 FIKEVSNADISMGDVATLSVTVIGIPKPKIQWFPNGVLLTPSADYKFPVFGDDHSLILIF 4288
QY 128 ----- 127
Db 4289 TKLEDEGEYTCMASNDYGKTCISAYLKINSKGEGKDTETESAVAKSLEKLGPCPPHPL 4348
QY 128 ----- 127
Db 4349 KELPIRCAQGLPAIFEYTVVGSBPAPVTWFKENKQLCTSVYVTI IHNPNGSGTFIVNDP 4408
QY 128 ----- 127
Db 4409 QREDSGLYICKAENMLGESTCAAEILLVLEDTMTDTPCKAKSTPEAPEDFPOTPLKGPA 4468
QY 128 -----YNHTAK----- 133
Db 4469 VEALDSEQEIATFVKDTILKAALITEENQOLS YEHIKAKANELSSQLPLGAQELQSILEQD 4528
QY 134 ----- 133
Db 4529 KLTPESTREPLCINGSIHFQPLKEPSNLQIQVSOQKTSKEGILMPEPETQAVLSDT 4588
QY 134 ----- 133
Db 4589 EKIFPSAMSIEQINSLTVEPLKTLIAEPGNYPOSSIEPPMHSYLTSAVEVLSPKEKTV 4648
QY 134 ----- 133
Db 4649 SDTNREQRVTLOKQEAQSALILISLSLAEGHVESLQSPDVMISQVNYEPLVPSHSCTEGG 4708
QY 134 ----- 133
Db 4709 KILIESANPLENAGDSAVRIEBGKSLRFPALAEKQVLLKEHSDNVVMPDQIIIESKR 4768
QY 134 ----- 133
Db 4769 EPVAIKKQVEQVRDILLSKESLLSGIPEEORLNKIQICRALQAASVSEQPLFSEWLRN 4828
QY 134 ----- 133
Db 4829 IEKVEEAVNITQEPHIMCMLVTSKSVTEETVIIIEDVDQPMANLKMELDALCAII 4888
QY 134 ----- 133
Db 4889 YEBIDILTAGPRIQOQAKTSLQEEMDSFGSQKVEPITEPEVESKYLISTERVSFNVQ 4948
QY 134 ----- 133
Db 4949 SRVKYLDATPVTKGASAVVDEKQDESUKPSBEKSESSESCTEAVTVKIQAEGLI 5008

QY 134 ----- 133
Db 5009 KEDGPMIHTPLDVTSEEGDIVHLTTSITNAKENVMYFENKLVPSDEKPKCLQDQNTYTL 5068
QY 134 ----- 133
Db 5069 VIDKVNTEHQGEYVCEALNDSKGTATSAKLTVVKRAAPVIRKRIEPLVALGHLAKFTC 5128
QY 134 ----- 133
Db 5129 EIQSAPNVRPQWPKAGREIYESDKSIRSSKYISSLEILRTQVVDCEYTCASNEYGSV 5188
QY 134 ----- 133
Db 5189 SCTATLTVTEAYPPTFLSRPKSLTTFVGAAKFICTVTGTPVIETIQDKGAALSPSPNW 5248
QY 134 ----- 133
Db 5249 RISDAENKHILELSNLTIQDRGVYSCASNKFGADICQAEIIIDKPHFKELEPVQSAI 5308
QY 134 ----- 133
Db 5309 NKKVHLECOVDEDRKVTVTWSDGQKLPPGKDYKICFEDKIATLEIPLAKLDSGTIVCT 5368
QY 134 -----TITPRPPP----- 141
 : : : : :
Db 5369 ASNEAGSSCSATVTVREPPPSFVKVDP SYMLPGESARLHCKLGSPVIQVTFWKNKE 5428
QY 142 ----- 141
Db 5429 LSESNPVMYFVNSEAILDITVVKVEDSGSYSCEAVNDVSGSDSCSTEIVIKEPPSFIKTL 5488
QY 142 ----- 141
Db 5489 EPADIVRGTNALLQCEVSGTGPEI SIFPKDKKQIRSSKKYRLFQKSLVCLEIFSPNSAD 5548
QY 142 -----CG----- 143
 : : : : :
Db 5549 VGEYECVAVANEVKGKCGCMATHLLKEPPTFVKVDDLIALLGGQTVTLQAAVRGSEPI SVTW 5608
QY 144 ----- 143
Db 5609 MKGQEVIREDGKIKMSFSGNVAVLIIIPDVOISFGGKYTCIAENEAGSQTSGSLIVKEPA 5668
QY 144 ----- 143
Db 5669 KIIERAELIQVTAGDPATLEYTVAGTPELKPWKYKDRPLVASKKYRISFKNNVAQLKPY 5728
QY 144 ----- 143
Db 5729 SAEIHDSCGYTPEISNEVSGSSCETTFTVLDRIAPFTKPLRNVDSVVNGTCRLDCKIA 5788
QY 144 ----- 143
Db 5789 GSLPMRVSWFKGCKEIAASDRYIAFVEGTASLEIIRVDMNDAGNFTCRATNSVGSKDS 5848
QY 144 ----- 143
Db 5849 GALLIVQEPATFVEKLEPSQLLKKGDATQLACKVTGTPPIKITWFANDREIKESSKHRMSF 5908
QY 144 ----- 143
Db 5909 VESTAVLRITDVGIEDSGEYMCQAQNEAGSDHCSIVIVKESPYFTKEFKPIEVLKEYDV 5968
QY 144 ----- 143
Db 5969 MLLAEVAGTPPEITWFKONTILRSGRKYKTFIQDHLVSLQILKFAADAGSYQCRVTNE 6028
QY 144 ----- 143
Db 6029 VGSSICSAVTLREPATITEEAVSIDVTOGDPATLQVKFSGTKETAKWFKDQBELTGS 6088
QY 144 ----- 143

Db	6089	KYKISVTDVTSILKIISTEKDQSGEYTFEVQNDVGRSSCKARINVLDLIIPPSFTKKLKK	6148	Db	7169	PELSVEWYKDGKLLTSSQKHKFSFYNKISSILRILSVRODAGTYTFQVQNNVGKSSCTAV	7228
QY	144	-----	143	QY	149	-----	148
Db	6149	MDSIKGSFIDLECIVAGSHPIQWPKDQDEISASEKYKFSFHDNTAFLEISQLEBPYPF	6208	Db	7229	VDVSDRAVPSPFTRRLKNTGGVLGASCILECKVAGSSPISVAVPHEKTKIVSGAKYQITTF	7288
QY	144	-----	143	QY	149	-----	148
Db	6209	VEKPOSQDVNPNTRVOLKALVGTA PMWTIKWFKDNKELHSGAARSVWKDDTSTSLELPAA	6268	Db	7289	SDNVCTQLNSLDSSDMGNYTCVAANVAGSDECRAVLTVQEPSPFVKPEPELVLPGKNV	7348
QY	144	-----	143	QY	149	-----	148
Db	6269	KATDSGTIQLSNDVGTATSKATLFVKEPPQFIKKPSPVLVRNQSTTFEQINGTPK	6328	Db	7349	TFTSVIRGTPPPKVNMPRGARELVKDRCNIFYEDTVALELFNIDI SOSGYTCVVSNN	7408
QY	144	-----	143	QY	149	-----	148
Db	6329	IRVSWYLDGNEITPAIOKHGIFSIDGLATFQISGARVENS GTYVCEARN DAGTASCSELK	6388	Db	7409	AGQASCTTREL FVKGLLKWIHPFCFLYTDILVFFLCITAVLCHLLBPAAFLKRLSDHSVEP	7468
QY	144	-----	143	QY	149	-----	160
Db	6389	VKEPPTPIRELKPEVVKISDVLECEVTGTPPFEVTWLKNNREIRSSKKYTLTDRVSVF	6448	Db	7469	GKSIILESTYTGTLPISVTWKKDGFNITTSKCNIVTTEKTCILBILNSTKRDAGQYSCE	7528
QY	144	-----	143	QY	161	-----	160
Db	6449	NLHITKCDPSDTGEYQICIVSNEGSCSCTRVALKEPPGFIKKIENTTTVLKSSATFQST	6508	Db	7529	IENEAGRDVCGALVSTLEPPYFVBLEPLEAAVGSVSLQCOVAGTPEITVSWYKGDTKL	7588
QY	144	-----	143	QY	161	-----	168
Db	6509	VAGSPISITWLKDDQILDDBDNVYISFVDSVATLQIRSVDNHSGRYTCQAKNESGVFR	6568	Db	7589	RTPPYRTYFTNNVATLVFNKVINDSGEYTCKAENSIGTASSKTIVFRIQERQLPPSPAR	7648
QY	144	-----	143	QY	169	-----	168
Db	6569	CYAFLLVQEPQAQIVKAKSVDVTEKDPMTLECVVAGTPELKVWLKDGKQIVPSRYFSMS	6628	Db	7649	QLKDIEQTVGLPVTLTLCRLNGSAPIQVCWYRQDGVLLRDDENLQTSFVDNVATLKILQTDL	7708
QY	144	-----	148	QY	169	-----	168
Db	6629	FENNVSFRIQSWKMGDSQGYTFKVENDFGSSCDAYLRVLDQNI PPSTFKLTKMDKVL	6688	Db	7709	SHSGYSCASNPLGTASSARLTAREPKSPFFDIKPVSIDVIAGESADFECHVTGAOP	7768
QY	149	-----	148	QY	169	-----	172
Db	6689	GSSIHECKVSGSLPI SAQWFKDGKEISTSAKRYLVCHERSVLSVNNLELEDTANYTCK	6748	Db	7769	MRIWTKDNKEIRPGGNYTITCVGNTPHLRILKVGKGDGQYTCQATNDVGDKMCSAQLS	7828
QY	149	-----	148	QY	173	-----	172
Db	6749	VSNVAGDDACSGILTVKEPPSFLVKPGRQQAIPDSTVEFKAILKGTPTPPKIKWFKDDVEL	6808	Db	7829	VKEPPKFKVBLEASKVAKQGESIQLECKISGPEIKVSWFRNDSSELHESWKYMSFINSV	7888
QY	149	-----	148	QY	173	-----	172
Db	6809	VSGPKCFIGLEGSTSFNLVYSVDASKTGQYTCHTVNDVGSDSCTTMLLVTEPPKFKVCKLE	6868	Db	7889	ALLTINEASAEDSGDYICEAHNGVGDASCSTALT VKAPPVFTQKPSPVGALKGSDVILQC	7948
QY	149	-----	148	QY	173	-----	172
Db	6869	ASKIVKAGDSSRLECKIAGSPEIRVVWFRNEHELPA SDKYRMTFIDSVAVIQNNILSTED	6928	Db	7949	EISGTPPPEVVVWKORKQVRSNKKFKITSKHFDTS LHLNLLEASDVGEYHCKATNEVGSD	8008
QY	149	-----	148	QY	173	-----	172
Db	6929	SGDFICEAQNPA GSTCSKTVI VKEPPVPSFPPIVETLKNAEVSLCELSGTPPEVVW	6988	Db	8009	TCSCSVKPEPPRFVKKLSDTSTLIGDAVELRAIVEGQFPI SVVWLKDRGEVIRESENTR	8068
QY	149	-----	148	QY	173	-----	172
Db	6989	YKDKRLRSSKKYKIASKNPHTSIHILNVDTSDIGEYHCKAQNEVGSDTCVCTVKLKEPP	7048	Db	8069	ISFIDNIATLQLGSPPEASNSGKYICOIKNDAGMRECSAVLTVLEPARIIEKPEPMTVTTCG	8128
QY	149	-----	148	QY	173	-----	172
Db	7049	RFVSKLSLTVVAGEPABLOASIEGAQPIFQVWLKEKEBEVIRSENI RITTFVENVATLQF	7108	Db	8129	NPPALECVVTGTPELSAKWFKDGRLELSADSKHHITFINKVASLKI PCAEMSDKGLYSFEV	8188
QY	149	-----	148	QY	173	-----	172
Db	7109	AKAEPANAGKYICOIKNDGGRENMATLWLEPAVIVEKAGPMTVTVTGETCTLECKVAGT	7168	Db	8189	KNSVGKSNCTSVVHSDRIVPPSFIRKLKDVNAILGASVLECRVSGSAPISVGWFOQGN	8248
QY	149	-----	148	QY	173	-----	172
				Db	8249	EIVSGPKCOSSFSSENVCTNLNLSLEPSDTGIYTCVAANVAGSDECSAVLTVQEPSPFEQT	8308

QY	173	-----	172
Db	8309	POSVEVLPGMSLTFTSVIRGTPPFVKWFKGRELVPGBSCNISLBDFVTELELFEVQPL	8368
QY	173	-----	172
Db	8369	ESGDYSCLVTNDAGSASCTHLLFVKBPATFVKRLADFSVETGSPVILEATYGTPTPISVS	8428
QY	173	-----	172
Db	8429	WIKDEYLIQSERCSTMTKSTILBILSTIEDYAQYSLIENEAGQDICEALVSVLEP	8488
QY	173	-----	172
Db	8489	PYPIEPLEHVEAVIGBPATLQCKVDGTPBIRISWYKHTKLSAPAYKQKNNVASLVI	8548
QY	173	-----	172
Db	8549	NKVDHSDVGEYSCKADNSGVASSAVLVIKARKLPPFPARKKDVHETLGFPPVAPECRI	8608
QY	173	-----	172
Db	8609	NGSEPLQVSWYKDGVLKKDANLQTSFVHVNATLQILOTDQSHIGQYNCASNPLGTASS	8668
QY	173	-----	172
Db	8669	SAKLILSEHEVPPFFDLKPVSVDLALGESGTFKCHVTGTAPIKITWAKDNREIRPGNYK	8728
QY	173	-----	172
Db	8729	MTLVENTATLTVLKVGKG DAGQVTCYASNIAGKDCSAQLGVQBPFRFIKLEPSRIVKQ	8788
QY	173	-----	172
Db	8789	DEFTRYECKIGSPEIKVLWYKDETEIQESSKFRMSFVDSVAVLEMHNLSVEDSGDYTC	8848
QY	173	-----	172
Db	8849	AHNAAGSASSTSLKVEPPIPRKKPHPIETLKGADVHLECELGQTPPFHVSWKDREL	8908
QY	173	-----	172
Db	8909	RSCKYKIMSENFLTSIHILNDAADIGBYQCKATNDVGS DTCVGSIALKAPPRFVKCLS	8968
QY	173	-----	172
Db	8969	DISTVVGKEVQLQTTTIEGABPISVVMFKGGEIVRES DNIWISYSENIA TLQPSRVEPAS	9028
QY	173	-----	172
Db	9029	AGKYTCQIKNDAGMQECFATLSVLEPATIVEKPESIKVTTGDTCTLECTVAGTPELSTKW	9088
QY	173	-----	172
Db	9089	FKDGKELTS DNKYKISFFNKVSGLKIINVAPSDGVSYPEVQNPVGKDSCTASLQVSDRT	9148
QY	173	-----	172
Db	9149	VPPSFRKLKETNGLSGSSVVMCKYVGPPIPSVSWFHEGNEISSGRKYQTTLTDTNICAL	9208
QY	173	-----	172
Db	9209	TVMLESDSDGYTCIATNAGSDECSAPLTVREPPSFVKQPDPMVDLGTNTVFTSIVK	9268
QY	173	-----	172
Db	9269	GTPPFSVWFKGSSSELVPGDRCNVLSLEDSVAELELFDVDTQSGEYTCIVSNEAGKASCT	9328
QY	173	-----TF-----	174
Db	9329	THLYIKAPAKFVRLNDYSIEKGKPLILEGTFGTGTPPISVTWKNGINVTSPQRCSNITTT	9388

QY	175	-----	174
Db	9389	EKSAILEIPSTVEDAGQYNCYENASGKDCSAQILILEQKLPPSFSRQLRDVQETVGL	9448
QY	175	-----	174
Db	9449	PVVPDCAISGSEPISVSWYKDGKPLKDSPNVQTSPLDNTATLNI PKTDRSLAGQYSC TAT	9508
QY	175	-----NPMFPNV-----	181
Db	9509	NFIGSASSARLILTEGKNPPFFDIRLAPVDVAVGESADFECHVTGTQPIKVSMAKDSRE	9568
QY	182	-----	181
Db	9569	IRSGGKYQISYLENSAHLTVLKVDKDSGQYTCYAVNEVGKDSCTAQLNIKERLIPPSFT	9628
QY	182	-----	181
Db	9629	KELSETVEBTEGNSFKLEGRVAGSQPITVAVYKNNIEIQPTSNCEITFKNNTLVLQVRKA	9688
QY	182	-----	181
Db	9689	GMNDAGLYTCKVSDNAGSALCTSSIVIKEPKPPVVDQHLTPVTVSEGEVQLSCHVOGS	9748
QY	182	-----	181
Db	9749	EPRIQWLKAGREIKPSDRCSFSPASGTAVLELRDVAKADSGDYVCASNAGSDTTKSK	9808
QY	182	-----	181
Db	9809	VTIKABPIQTKRIQNI VVSEHQSATFECEVSFDDAIVTWYKPTTELTSQKNFRNDGR	9868
QY	182	-----	181
Db	9869	CHYMTIHNVTPDDEGVYSVIARLEPRGEARSTAE LYTLTKEIKLELKPDPIDPSRVP IPT	9928
QY	182	-----	181
Db	9929	MPIRAVPPEEIPPVVAPPILLLPTPEKKPPPKRIEESHERKVPKAKVPEKKAPPPPKVI	9988
QY	182	-----	181
Db	9989	KKPVIEKIEKTSRRMEEEKVQTVKPEIKPAIPLPAPEPKPKEAGVPKKTPTSPIAEARR	10048
QY	182	-----	181
Db	10049	KLRPGSGKEPPDEAPFTYQLKAVPLKFVKEIKDILITSESEFVGSSAIFEC LVSPTAIT	10108
QY	182	-----	181
Db	10109	TWMDGSNIRESPKHFITADGDKRLHIIDVOLSDAGEYTCVLR LGNKEKTSTAKLVVEE	10168
QY	182	-----	181
Db	10169	LPVRFVLTLEEBVTVKGOPLYLSCELNKERDVWRKOGKI VVEKGRIVPGVIGLMRAL	10228
QY	182	-----	181
Db	10229	TINDADDDTAGTYTTVTENANNLECSSCVKVVVEIRDWLVPIRDQHV KPGTAFACDI	10288
QY	182	-----	181
Db	10289	AKDTPNIKWFGYDEIPABPNDKTEILRDGNHLYLKIKNAMPEDIAEYAVEIEGKRYPAK	10348
QY	182	-----	181
Db	10349	LTLGEREVELLKPIEDVTIYKESAFDAEISEADIPGQWK LKGLLRSPSTCEIKAEKG	10408
QY	182	-----	181
Db	10409	KRFLTLHKVLDQAGEVLQALNAITTAITLTVKEIELDFAVPLKDVTVPRROARFECVL	10468
QY	182	-----	181

QY	182	-----	181	QY	193	-----TKNVND-----	198
Db	12689	ILKQDKRITINVPKSTVTIIVDSKRSDTGTIIIEAVNVCGRATAVVEVNVLDKGPAA	12748	Db	13769	IYEDVTSNMLYKWNPKDNGSPIILGYWLEKREVNSTHWSRVNKSLLNALKANVDGLLEG	13828
QY	182	-----	181	QY	199	-----SOTI-----	202
Db	12749	FDITDVNESCLLTWNPPRDDGSKITNVYVERRATDSEVWHKLSVTSKDTNFKATKLIP	12808	Db	13829	LTYVFRVCAENAAAGPKSPSPDPKTAHDPIISPFGPPPIPRVTDTSSTTIELEWEPPAFNG	13888
QY	182	-----	181	QY	203	-----YFL-----	205
Db	12809	NKEYIFRVAENMYGVEPVQASPIITAKYQFDPGPPPTRLPSDITKDAVTLTWCEPDD	12868	Db	13889	GGEIVGYFVDKQLVGTNEWSRCTERKMIKVRQYTVKEIREGADYKLRVSAVNAAGEPPGE	13948
QY	182	-----	181	QY	206	-----	205
Db	12869	GGSPITGYWVERLDPTDKWRCNKMFKVDTTYRVKGLTNKKYRPRVLAENLAGPKPS	12928	Db	13949	TQPTVAEPQEPAPVELDVSVKGGIQIMAGKTLRIPAVVTGRPVPTKVTWKBEGELDKOR	14008
QY	182	-----NT-----	186	QY	206	-----	205
Db	12929	KSTEPILIKDPIDPPSPRRLWLEVINITKNTADLKWTVPEKGGSPITNVIYVEKRDVRRKG	12988	Db	14009	VVIDNVGTKSELLIKDALRKDHGRVITATNCGSKGFAAARVEVPDVGPFVLDLKPVVTN	14068
QY	187	-----KLJV-----	190	QY	206	-----GLTALL-----	211
Db	12989	WQTVDTTVKDKCTVTPLTEGSLYFRVAENAIQSDYTEIEDSVLAKDTFTTPOPPYA	13048	Db	14069	RKWCLLNMSDPEDDGGSEITGFIIEBKDAKMTWRQPIETERSKCDITGLLEGQEKYFRV	14128
QY	191	-----	190	QY	212	-----LRYAQR-----	217
Db	13049	LAVVDVTKRVDLKWBPKNKGGRPIQRYVIEKKERLGRWVKAGKTAGDCNFRVTDVI	13108	Db	14129	IAKNKFGCGPPVEIGFILAVIDPLGPPTSPERLTYTERTKSTITLDMKEPRNGSGSIOGY	14188
QY	191	-----	190	QY	218	-----	217
Db	13109	EGTEVOQVRAENAGVHPSEPTIELSDTPSPPLDLHVTDAGRKHIAIAWKPE	13168	Db	14189	IIIEKRHDKDPFERVKNKLCPTTSFLVENLDEHQMVEFRKAVNEIGSEPSLPLNVVIQ	14248
QY	191	-----	190	QY	218	-----	217
Db	13169	KNGGPIIGHVHEMCPGTEKWNVNSRPIKOLKPKVEGVPVDEKYVLRVAVNAIGVS	13228	Db	14249	DDEVPTIKLRLSVRGDTIKVKAGBPVHIPADVTGLPMPKIEWSKNETVIEKPTDALQIT	14308
QY	191	-----	190	QY	218	-----	217
Db	13229	EPSEISENVAKDPCKPTIDLETHDIIIEGEBKLSIPVFRAPVPTVSWHKDGKEVKA	13288	Db	14309	KEEVSERSEAKTELSIPKAVREDKGTVTYTASNRLGVSFRNVHVVYDRSPRNLAVTDI	14368
QY	191	-----	190	QY	218	-----	217
Db	13289	SORLTWKNNDHISAHLEVPKSVRADAGIYITILENKLGSATASINVKIVGLPGCKDIKAS	13348	Db	14369	KAESCYLTDAPLDNGGSEITHYVIDKRDASRKKAEBEVTNTAVEKRYGIWKLIPNGQY	14428
QY	191	-----	190	QY	218	-----	217
Db	13349	DITKSSCKLTWBPPEFGGTPIILHYVLRERAGRTYIIPVMSGENKLSWTVKDLIPNGEY	13408	Db	14429	EFRVAVNKYGISDECKSKVVIQDPYRLPGPPGPKVLARTKGSMLVSWTPELDNGGSP	14488
QY	191	-----	190	QY	218	-----	217
Db	13409	FPRVAVNKVGGEYIELKNPVIAQDKPPDPVDVEHNPTAEAMTITWKPLDYDGS	13468	Db	14489	ITGYWLEKREEGSPYWSRVSRAPITKVGLKGVEFNVPRLLLEGVKYQFRAMAINAAGIGPP	14548
QY	191	-----	190	QY	218	-----	217
Db	13469	KIMGYIIIEKIAKEERWKRNEHLVPIITYTAKLEEGKEYQPRVRAENAAAGISEPSRAT	13528	Db	14549	SEPSDEVAGDPIFPFGPPSPCEVKDKTKSSISLWKPKPAKDGGSPIKGIYIEMQEGYT	14608
QY	191	-----	190	QY	218	-----	217
Db	13529	PPTKAVDPIDAPKVIILRTSLEVKRGDEIALDASISGSPYTTIWKDENVIVPEEIKKRA	13588	Db	14609	DMKRVNEPDKLITTCCECVVPNLKLRLKYRPRVKAVNEAGESEPSDTTGEIPATDIQEEPE	14668
QY	191	-----	190	QY	218	-----	217
Db	13589	APLVRERKEVQEEFPVLPLTQRLSDNSKKGESQLRVDRDSLRPDHLGYMIKVENDHGI	13648	Db	14669	VFIDIGADCLVKAGSQIRIPAVIKGRPTPKSSWEFDGKAKMKLETAENSVIIIBE	14728
QY	191	-----	190	QY	218	-----NC-----	219
Db	13649	AKAPCTVSLDTPGPPINFVEDIRKTSVLCKWEPLDDGGSEIINYTLEKKDKTKPDSE	13708	Db	14729	CKRSHTGKYSITAKNAGOKTANCRVKVMDVPGPKDLKVSDITRGSCLSWMPDDGG	14788
QY	191	-----GP-----	192	QY	220	-----	219
Db	13709	WIVVTSLRHCKYSVTKLIEGKEYLPRVRAENRFGPPPCVSKPLVAKOPFGPPADAPDKP	13768	Db	14789	DRIKGVIEKRTIDGKAWTKVNPDCGSTTFVVPDLLSEQQYFFRVRPAENRFGIGPPVETI	14848
				QY	220	-----THSFYLVNAM-----	229

Db	14849	QRTTARDEIYPPDPPIKLIKIGLITKNTVHLSWKPKNDGGSPVTH--YIVECLAWDPTGT	14906
Qy	230	-----	229
Db	14907	KKEAWROCNKRDVEELOFTVEDLVEGGEYEFVRKAVNAAGVSKPSATVGPVTVKDOQCPP	14966
Qy	230	-----	229
Db	14967	SIDLKEFMEVEEGTDVNI VAKI KGVPPTLTWFKAPPKPKDNKEPVLVDTHVKNLVDDT	15026
Qy	230	-----	229
Db	15027	CTLVIQSRRSRDTGLYITITAVNNLGTASKEMRLNVLGRPPVGPPIKFESVSADOMTSLW	15086
Qy	230	-----	229
Db	15087	FPKDDGSKITNVIIEKREANRKTWVHSSEPKECTYTTPKLLGHEHYFRIMAQNKYG	15146
Qy	230	-----SRNLFVRP :	237
Db	15147	IGEPDSEPETARNLFSVPGAPDKPTVSSVTRNSMTVNWBEPEYDGGSPVTGYWLEMKDT	15206
Qy	238	-----	237
Db	15207	TSKRWRKVRNRPDKAMTLGVSXKVTGLIEGSDYQFRVYAINAAGVGPASLPSPATARDP	15266
Qy	238	-----KYINGTKL--	245
Db	15267	IVTDWTKSSADLEWSPLKDGGSKVGYIIVEYKEEGKEWEKDKVEVRGTGLVVTGLKE	15326
Qy	246	-----	245
Db	15327	GAFYKFRVRAVNIAGIEPGEVTDVIEBKORLVSPLQLDASVRDRIVWHAGGVIRIAY	15386
Qy	246	-----	245
Db	15387	VSGKPPPTVWNMERTLPOEATIEATISSMWIKNCQSHQGVYSLAKNEAGERKKT	15446
Qy	246	-----KNT--	248
Db	15447	IIVDVLAVPERPEDLEVKVTKNTVTLTWNPVKYDGGSEIINYLBSRLIGTEFKHKVTN	15506
Qy	249	-----	248
Db	15507	DNLLSRKYTVKGLKEGDTYEVRSVAVNIVGQKPSFCTKPIITCKBELDKLITIRVGQAFAL	15566
Qy	249	-----MRKJKEKQ--	256
Db	15567	TGRYSGKPKPKVSWFKDEADVLEDDRTHIKTTPATLALALEKIKAKRSDSGKYCVVVENSTG	15626
Qy	257	-----	256
Db	15627	SRKGFQVNVVDHPGPPVGPVSFDEVTKDVMVLSWKPPPLDDGSKITNYIIEKKEVGKV	15686
Qy	257	-----	256
Db	15687	WMPVTSASAKTCKVSKLLEGKDYI FRIHAENLYGISDPLVSDSMKAKORFRVPDAPDQ	15746
Qy	257	-----	256
Db	15747	IVTEVTKDSALVTWPKHDGKPIITNYIIEKRETMKRWARTKDPHPTTKFRVPDILLE	15806
Qy	257	-----	256
Db	15807	GCQYEFVRAENEIGIDPSPSKPVFAKDPFAKPSPPVNPPEAIDTCNSVDLTWQPPRH	15866
Qy	257	-----	256
Db	15867	DGGSKILGYIIVEYQKVGDSEWRANHTPSCPETKYKVTGLRDGQTYKFRVLAVNAAGES	15926
Qy	257	-----	256

Db	15927	DPAHVPEPVLVKDRLEBPPELLIDANNAREQHIVGDTLRLSAIIKGVPPPKVTWKEDRD	15986
Qy	257	AP-----	258
Db	15987	APTAKRIDVTPVGSKLEIRNAAHEDGGIYSLTVENPAGSKTVSVKVLVDKPGPPRDLEV	16046
Qy	259	-----	258
Db	16047	SEIRKDSCLYTWKEPLDDGGSVITVYVERRDVASAQWSPLSATSKKSHFPAKHLNEGNO	16106
Qy	259	-----	258
Db	16107	YLFRAAENQYGRGPPVETPKPIKALDPLHPGPKDLHHVDVDKTEVSLVWNKPRDGG	16166
Qy	259	-----	258
Db	16167	SPITGYLVBEGTODWIKFTVTNLCVVTGLQQGKTYRFRVKAENIVLGLSPDTTIP	16226
Qy	259	-----	258
Db	16227	IECQEKLVPPSVVELDVKLI EGLVVKAGTTRPFAIIRGVVPFAKWTTDGSEIKTDEHYT	16286
Qy	259	-----	258
Db	16287	VETDNFSSVLTIKNCLRRDTGEYQITVSNAGSKTVAVHLTVLDPGPTGPINILDVTP	16346
Qy	259	-----	258
Db	16347	EHMTISWQPPKDDGGSVINYIVEKQDTRKOTWGVVSGSSKTKLKI PHLQGEYVFRV	16406
Qy	259	-----	258
Db	16407	RAENKIGVPLDSTPTVAKHKFSPSPPGPKPVVTDITENAATVSWTLPKSDGSGSPITY	16466
Qy	259	-----	258
Db	16467	YMERREVTGKVRVVKNTPIADLKFRVTGLYEGNTYEFVFAENLAGLSKSPSSDPKAK	16526
Qy	259	-----	258
Db	16527	RPKPPGPPINPKLKDKSRETADLVWTKPLSDGSGPILGYVVECKPGTAQWRINKDEL	16586
Qy	259	-----	258
Db	16587	IRQCAFVPLIEGNEYFRFRIKAANI VGEGEPELAESVIAKDILHHPPEVELDVTCDVI	16646
Qy	259	-----	258
Db	16647	TVRVQOTIRILARVKRPEPDI TWTKGKVLVREKRVDLIQDLPRVELQI KEAVRADHGK	16706
Qy	259	-----	258
Db	16707	YIISAKNSSGHAQSAI VNVLDRLPGCQNLKVTVTKENTISWENPLDNGSGSEITNFIV	16766
Qy	259	-----	258
Db	16767	EYRKPNQGSIVASDVTKRLIKANLLANNEYFRVCAENKVGVPGTIETKTPILAINPI	16826
Qy	259	-----	258
Db	16827	DRPGEPENLHIADKGTTFVYLKWRPPDYDGGSPNLSYHVERRLKSGDDWVERVHKSGIKET	16886
Qy	259	-----VKEQFBK-- :	265
Db	16887	HYMDRCVENQIYEFVQTKNEGGSWVKTEEVVVVKEDLQKPVLDLKL SGLVTVKAGDT	16946
Qy	266	-----KAK-- 	268
Db	16947	IRLEAGVRGKPPPEVAVWTKDADTLTRS PRVKIDTRADSSKFSILTKAKRSDGGKYVVA	17006
Qy	269	-----	268
Db	17007	TNTAGSFVAYATVNVLDKPGPVRNLKIVDVSSDRCTVCWDPDDGGCEIQNYILEKCE	17066

QY	269	-----	268	QY	290	-----	289
Db	17067	KRMVSTYSATVLTGTTVRLIEGNEYIFRVAENKIGTPPTESKPVIAKTKYDKPGR	17126	Db	18147	GLQEGTEYFRVTAINKAGPKSPDASKAAYARDPQYPPGPPAPPKYVDTRSSVLSWG	18206
QY	269	-----	268	QY	290	-----	289
Db	17127	PDPEVTKVSKEEMTVVWNPPEYDGGKSIITGYFLEKKEKSHTRWVPVNSAIPERRMKVQ	17186	Db	18207	KPAYDGGSPIIIGYLVEVKADSDNWVRCNLQNLQKTRFEVTLGMDTOYQPRVAVANKI	18266
QY	269	-----	268	QY	290	-----	289
Db	17187	NLLPDHEYQPRVKAENEIGIBSPSLPSRPVAKDPIEPGPPPTNFRVVDTTKHSITLGMG	17246	Db	18267	GYSDDSDVPDKHVPKDIILIPPEGELDADLRKTLILRAGVTMRLYVPVKGRPPPKITWSKP	18326
QY	269	-----	268	QY	290	-----	289
Db	17247	KPVYDGGAPIIGYVVENRPKIADASPDEGMRKRNAAQAALVRKEFTVTSLDENOEYFRMS	17306	Db	18327	NVNLDRIGLDIKSTDFTFLRCENVNKYDAGKYLITLNSCGKKEYTIVVKVLDTPGPP	18386
QY	269	-----	276	QY	290	TNVT-----	293
Db	17307	LKHGHVSNAPPENDGSGQVTHIVEKREADRTWSTVTEVKKTSFHVTNLVPGNYYF	17366	Db	18387	VNVTVKEISKDSAYVTWEPPIIDGGSPIIINYVVKQDAERKSHSTVTTECSKTSFRVPNL	18446
QY	277	-----	276	QY	294	-----	293
Db	17367	RVTAVNEYGPVTDVPKPVLASDPLSEDPDRKLEVTMTKNSATLAWLPLPLDGGAKI	17426	Db	18447	EEGKSYFRVFAENEYIGDGPGETRDAVKASQTPGPVVDLKVRSVSKSCSIGWKKPHSD	18506
QY	277	-----	276	QY	294	-----	297
Db	17427	DGYITSYREBEQPADRWTEYSVVVDLSLVVTGLKEGKVKFRVAARNAVGVSIPREAEV	17486	Db	18507	GGSRIGYVVDFTLTKENKQVNMKSLQYSAKDLTEGKEYTFRVSAENENGEGTPTSEIT	18566
QY	277	-----	276	QY	298	TAAR-----	301
Db	17487	YEAKEQLLPKILMPQIITIKAGKURIEAHVYGVKPHPTCKWKKGDEVVVTSHLAVHKA	17546	Db	18567	VVARDDVAPDLDLKGLPDLCYLAKENSFRLKPIKIGKPAVSVWKGEDPLATDTRVS	18626
QY	277	-----	281	QY	302	-----	301
Db	17547	DSSSILIIKDVTBKDSYSLTAENSSGTDQKI KVVVMDAPGPPQPPDISIDADACS	17606	Db	18627	VESSAVNTTILVYDCQKSDAGKYITILKNVAGTKEGTISIKVVVGKPIPTGPIKFDVTA	18686
QY	282	-----	281	QY	302	-----	301
Db	17607	LSWHIPLEGGSNITNYIVEKCDVSRGDWVTALASVTKTSRKGKLIPOGEYIFRVAEN	17666	Db	18687	EAMTLKWAAPPKDDGGSEITNYILEKRDVNNKVVTCASAVQKTTFRVTRLHEGMEYTFRV	18746
QY	282	-----	281	QY	302	-----	301
Db	17667	RFGISBLTSPKMAVQPPFGVPSEPKNARVTKNKDCIFVWDRPDSOGGSPIIIGYLIER	17726	Db	18747	SAENKVGEGGLKSEPIVARHFPDVPDAPPPNIVDVRHDSVSLTWDPKTKGGSPIITGY	18806
QY	282	-----	281	QY	302	-----	301
Db	17727	KERNLLVWKANDTLVRLTEYPCAGLVEGLEYSFRIYALNKAGSSPPSKPTEYVTARMPV	17786	Db	18807	HLEFKERNLLMKRANKTPIRMDFKVTGLTEGLEYEFRVMAINLAGVGKPSLSPBPVVA	18866
QY	282	-----	281	QY	302	-----	301
Db	17787	DPGKPEVIDVTKSTVLIWARKPHDGGSKIIGYFVEACKLPGDKWRCNTAPHQIPQEE	17846	Db	18867	LDPLEAPTIVLDPTIKDGLTIKAGDTIVLNAISILGKPLPKSSWSKAGKDIRPSDITQIT	18926
QY	282	-----	281	QY	302	-----	301
Db	17847	YATGLEKAQYQFRAIARTAVNISPPSESDPVTILAENVPPRIDLSVAMKSLLTVKAG	17906	Db	18927	STPTSSMLTIKYATRKADAGEYITITATNPFGTKVEHVKVTVLDPVPGPPVPEISNVSAEKA	18986
QY	282	-----	289	QY	302	-----	304
Db	17907	TNVCLDATVPGKPMPTVSWKKGDTLLKPAEGIKMAQORNLCITLLEFSVNRKDSGDYITA	17966	Db	18987	TLTWTTPLEDGGSPISKYILEKRETSRLLTWTVVSEDIQSCRHVATKLIQGNEYIFRVS AV	19046
QY	290	-----	289	QY	305	-----	312
Db	17967	ENSSGSKSATIKLVLDKPGPPASVKINKMYSRAMLSWEPPLEDGGSEITNYIVDKRET	18026	Db	19047	NHYGKEPVQSEFPVKWVDRFGPPPEKPEVSNVTNTATVSMKREVDGSGSEITGYHVE	19106
QY	290	-----	289	QY	313	-----	312
Db	18027	SRPNWAQVSATVITSCSVEKLIIEGHEYQFRIKAENKYGVDVFTPEPAIAKNPYDPPGR	18086	Db	19107	RREKSLRWVRAIKTPVSLRCKVTGLQEGSYEFVRVSAENRAGIOPPPPPGPPSNPHVT	19166
QY	290	-----	289	QY	313	-----	326
Db	18087	CDPPVISNITKDHMTVSWKPPADDDGSGPITGYLLEKRETAQVNWTKVNRKPIIERTLKAT	18146	Db	19167	DTTKKSASLAWKGPYDGGLEITGYVVEHQVKGDEAWIKDITCTALRITQFVVVPDLQTKE	19226
				QY	327	-----	326

Db	19227	KYNFRISAINDAGVGEPAVIPDVEIVEREMAPDFELDAELRRTLVRAGLSIRIFVPIKG	19286	Db	20307	CCTURLFVPIKGRPAPEVKWARDHGESLDKASIBESTSYTLLIVGNVRPDSGKYILTVE	20366
Qy	327	-----	326	Qy	340	-----	339
Db	19287	RPAPVETWTKDNINLKNRANIENSTFLLIIPECNRYDTGKFMVTIENPAKKSGFVNV	19346	Db	20367	NSSGSKSAFVNVRLDTPGPPQDLKVKVEVTKTSVTLTWDPLLDGSGSKINYIVEKRE	20426
Qy	327	-----	326	Qy	340	-----	339
Db	19347	RVLDTGCVLNLRPTDITKDSVTLHWDLPLIDGSRITNYIVEKREATKSYSTATTKCH	19406	Db	20427	RKAYSTVATNCHTKSWKVDLOEGCSYYFRVLAENEYIGILPAETAESVKASERPLPPGK	20486
Qy	327	-----	326	Qy	340	-----	339
Db	19407	KCTYKVTGLSEGCEYFRVMAENEYIGIBPTETTEPVKASEAPSPDLSLIMDITKSTVS	19466	Db	20487	ITLMDVTRNSVLSWEKPEHGGSRILGYIVEMQTKGSDKWATCATVKVTEATITGLIQG	20546
Qy	327	-----	326	Qy	340	-----	339
Db	19467	LAWPKPHDGGSKITGYVIEAQRKSGDQWTHITTVKGLCEVVRNLTEGSEYTFQVMVNS	19526	Db	20547	BEYSFRVSAQNEKGISDPRQLSVFVIAKDLVIPPAFKLLFNTFTVLAGEDLKVDPFIFGR	20606
Qy	327	-----	326	Qy	340	-----	339
Db	19527	AGRSAPRESRPVIVKESQMLPELDLRLGIYQKLVIAKAGDNIKVEIPVLGRPKPTVTKKG	19586	Db	20607	PTPAVTWHKDNVPLKQTTTRVNAESTENNSSLITIKDACREDVGHVYVVKUTNSAGEAIE	20666
Qy	327	-----	326	Qy	340	-----	339
Db	19587	DQILKQTVRVPETTATSTILNINECVSRSDGFPYPLTARNIVGEVGDVITIQVHDIPGPP	19646	Db	20667	VIVLDKPGPTGPVKMDEVTAADISITLSWGPVKYDGGSSINNYIVEKRDTSITTTQIVSAT	20726
Qy	327	-----	332	Qy	340	-----	339
Db	19647	TGPIKFDEVSSDFVTFSWDPPENDGGVPISNYVEMRQDSTTVELATTVIRTTTKATR	19706	Db	20727	VARTTIKACRLTKGCEYQFRIAENRYGKSTYLNSEPTVAQYVPFKVPGPGTVPVTLSSR	20786
Qy	333	-----	332	Qy	340	-----	339
Db	19707	LTTGLEVQFRVKAQNRGVGPGITSACIVANYFPKVPGPPTQRIPGYRARIAYEVRVI	19766	Db	20787	DSMEVQWNEPISDGGSRVIGYHLECKERNISILWVKLNTPIQTKFKTGLLEGVEYEF	20846
Qy	333	-----	332	Qy	340	-----	339
Db	19767	AENMAGSKSPSEPMALDPIDPPGKVPPLNITRHTVTLKWAKPEYTGFGKITSYIVE	19826	Db	20847	VSAENIVGIGKPSYSECYVARDPCDPGRPEALIVTRNSVTLQWKPTYDGGSKITGYI	20906
Qy	333	-----	332	Db	20907	VEKKELPEGRWMKASFTNIIDTHFEVTGLVEDHRYEFVRVIARNAAGVSEPSSESTG	20966
Db	19827	KRDLNGRWLKANFSNILENEFTVSGLTEDAAYEFVRVIAKNAAGAISSPSEPSDAITCRD	19886	Qy	340	-----	339
Qy	333	-----	332	Db	20967	RDEVDPRIISMDPKYKTIIVHAGESFKVDADIYKPIPTIOWIKGDQELSNARLEIKS	21026
Db	19887	DVEAPKIKVDKFDTVILKAGEAFRLEADVSGRPPPTMEWSKDGELEGTAKLEIKIAD	19946	Qy	340	-----	339
Qy	333	-----	332	Db	21027	TDFATSLSVKDAVRVDSGNYILKAKNVAGERSVTNVKVLDRPDPGPEGVVISGVTAEC	21086
Db	19947	FSTNLVNKDSRDSGAYTLTATNPGGFAKHFNVKVLDRPDPGPEGLAVTEVTSEKCVL	20006	Qy	340	-----	339
Qy	333	-----	332	Db	21087	TLAWKPLODGGSDIINYIVERRETSRLVWTVVDANVQTLSCVKTLLEGNEYTFRIMAV	21146
Db	20007	SWFPPLDDGGAKDHYIVQKRETSRLAWTNVASEVQTKLVTKLKGNEYIFRVMVANK	20066	Qy	340	-----	343
Qy	333	-----	332	Db	21147	NKVGCEPLESEPVAKNPFVVPDAPKAPVTTVTKDSMIVVERPASDGGSEILGYVLE	21206
Db	20067	YGVGCEPLESVLAVNPYGPDPKPNPEVTTITKDSMVVCMWGHPSDGGSEIINYIVER	20126	Qy	344	-----	346
Qy	333	-----	339	Db	21207	KRDKEGIRWTRCHKRLIGELRLRVLTGLIENHDYEFVRSNAAGLSEPPSPAYQKACDP	21266
Db	20127	DKAQRWIKCNKKTLDLRYKVSLTEGHEYEFPRIMAENAGISAPSFTSFYKACDTVF	20186	Qy	347	-----	346
Qy	340	-----	339	Db	21267	IYKPGPNPNPKVIDITRSSVFLSWSKPIYDGGCEIQGYIVEKCDVSGVGTWCTPPTGIN	21326
Db	20187	KPGPPGNPRVLDTSRSSISIAWNKPIYDGGSEITGYMVETALPEDEWQIVTTPAGLKAT	20246	Qy	347	-----	346
Qy	340	-----	339	Db	21327	KTNIEVKLEKHEYNFRICAINKAGVGEHADVPGPPIIVEEKLEAPDIDLDLELRKIINI	21386
Db	20247	SYTTITLTENQYKIRIYAMNSEGLGBPALVPGTFKAEADRLMPPPEIELDADLRKVVTIRA	20306	Qy	347	-----	346
Qy	340	-----	339	Db	21387	RAGGSURLFVPIKGRPTPEVKMGKVDGEIRDAAIIDVTSSFTSLVDNVRVDSGKYTLT	21446

QY	347	----	LYY-	381	----	LYY-	383
Db	21447	LENSGKSAFVTVRLDTPSPVNLKVTETKDSVITWEPPLDGGSKIKNIYVEKRE	21506	Db	22527	NLTEGAIYYFRVMAENEFVGVPVETVDAVKAAPPPGKVTLTVDVQSOTASILMWEKPE	22586
QY	347	----		QY	384	----	383
Db	21507	ATRKSAVVVNTCHNSWKIDQLQEGCSYYFRVTAENEYIGLPAQTADPIKVAEVPQPP	21566	Db	22587	HGGSRVLGVVEMQPKTEKWSIABSKVCNAVVTGLSSGQYQFRVKAINEKGSDDR	22646
QY	347	-----SNRTAVS-	354	QY	384	----	383
Db	21567	GKTVDDVTNSVLSWTKEPHDGGKIIQYIIVEMQAKHSEKSECARVKSLOAVITNLT	21626	Db	22647	VLGVPVIAKDLTIQPSLKLFPNTYSIQAGEDLKIEIPVIGRPPNLSWKDGEPLKQTR	22706
QY	355	----EFM-	357	QY	384	----	383
Db	21627	QGBEYLFVVAVNEKGRSDPRSLAVPIAKDLVIEPDVKPAPSSYSVQVQDLKIEVPIS	21686	Db	22707	VNVEETATSTVLHIKEGNKDDFKVTVTATNSAGTATENLSVILEKPGPPVGPVRFDEV	22766
QY	358	-----KNTH-	361	QY	384	----	383
Db	21687	GRPKPTITWKGPLKQTRINVTDSLDTLTSIKETHKDDGGQYITVANVVGOKTAS	21746	Db	22767	SADFFVISWEPPAYTGGCOISNYIVEKRDTTTTTHMVSATVARTTIKITKLKTGYQF	22826
QY	362	----	361	QY	384	----	383
Db	21747	IEIVTLKDDPPKGPVKFDDVSAESITLSWNPLYTGGCOITNYIVQKRDTTTTVWDVVS	21806	Db	22827	RIFAENRYKSAPLDSKAVIQVPPKEPGPGTFFVTSISKQMLVQWHEPVNDGGTKII	22886
QY	362	----	361	QY	384	-----NET-	386
Db	21807	ATVARTTLVKTKTGYQFRIFAENRYGQSFALSDPIVAQYPYKEPGPGKASKNSE	21866	Db	22887	GYHLEQEKNSILWVKLNKTPIQDTKFTKTGLDEGLEYEFKVSANIIGIGKPSKVSECF	22946
QY	362	----	361	QY	387	----	386
Db	21867	CYVARDPCDPGTPPEIPMVKRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRWMKASFT	21926	Db	22947	VARDPCDPPRPEAIVITRNNTLVKWKPAYDGGSKITGYIVEKDLDPGRWMKASFTNV	23006
QY	362	----	361	QY	387	----	386
Db	21927	NVETQTVSGLTDORYEFRVIKNAAGAKSPDSSTGPITAKDEVLPRIEMDPKFRD	21986	Db	23007	LETEFTVSLVEDQRYEFRVIAARNAAGNFSEPSDSSGATARDEIDAPNASLDPKYKQVI	23066
QY	362	----	361	QY	387	----	386
Db	21987	TIVNAGETFRLEADVHGKPLPTIEWLRGDKEIESARCEIKNTDPKALLIVKDAIRIDG	22046	Db	23067	VVHAGETFVLEADIRGKPIPDVVMWSDGKELEBETARMBIKSTIQKTTLLVVKDCIRTDGG	23126
QY	362	----	361	QY	387	----	386
Db	22047	GQYILRASNAVAGSPFVNVKLDLRPGPPGVPQVTGVTSEKSLTWSPLQDGGSDISH	22106	Db	23127	QYILKLSNVGKTSIPITVKVLDLRPGPPGPKLVGTGVTABKCYLAWNPPLQDGGANISHY	23186
QY	362	-----VLIR 365	22106	QY	387	----	386
Db	22107	YVVEKRETSRLAWTVVASEVVTNSLKVTKLLEGNEYVFRIMAVNKYGVGEPLSAPVLMK	22166	Db	23187	IIKRETSRLSWTQVSTEQALNYKVTKLLPGNEYIFRVMVNVKYGIGEPLESGPVTACN	23246
QY	366	N-----	366	QY	387	----	386
Db	22167	NPVLPGPSLEVTNIAKDSMTVCWNRDSDGSEIIGYIVEKDRSGIRWIKCNKRI	22226	Db	23247	PKPPGPPSTPEVSATTKDSMVVTWARPVDDGTEIEGVILEKRDKEGVRTWKCNKTLT	23306
QY	367	-----ETPYTIY-	373	QY	387	----	386
Db	22227	TDLRLRVTLGTEDHEHYEFRVAENAGVGEPSATVYKACDPVFKPGPTNAHIVDTTK	22286	Db	23307	DLRLRVTLGTGHSYEFRAENAGVGEPSBPSVFRACDALYPPGPPSNPKVTDTSRS	23366
QY	374	----	373	QY	387	----	386
Db	22287	NSITLAWKPIYDGGSEILGYVVEICKADEEHWIVTPQGLRVTRFRIKSLTEHQEYKI	22346	Db	23367	SVSLAWSKEIYDGGAPVKGYVVEVEKAAADEWTTCTPPTGLOKQFTVTYKLENTENYFR	23426
QY	374	-----GT-	380	QY	387	----	386
Db	22347	RVCALNKVGLGEATSPGTVKPEDKLEAPELDDSELRGIVVRAGGSARIHIPKGRPT	22406	Db	23427	ICAINSEGVGEPATLPGSVVAQERIEPPEILDADLRKVVLASATLRLFVTIKGRPEP	23486
QY	381	----	380	QY	387	----	386
Db	22407	PEITWSREBEFTDKVQIEKGVNYTQLSIDNCDNRNDAGKYILKLENSGSKSAFVTVKVL	22466	Db	23487	EVKWEAKGILTDRAGIEVTSFTMLVIDNTRFDSGRYNTLLENNSSKTAFAVNVRLD	23546
QY	381	----	380	QY	387	----	386
Db	22467	DTPGPPONLAVKEVRKDSAFVWEPPSLMEQGRSSNYVIDKPKAYANVSSKCKTSFKVE	22526	Db	23547	SPSAPVNLTIREVKKDSVTLSEPPPLIDGGAKITNYIVEKRETRKAYATIINNCTKTTTF	23606
				QY	387	----	386

Db	23607	RIENLOGCSYYFVRLASNEYIGLPAETTEPVKVSEPLPGRGRTLVDVTRTATIKWE	23666	Db	24687	WOQPAFDGSGKITGVIVERRDLPDGRWTKASTNTVTETQFIISGLTQNSQYERFVAFARNA	24746
QY	387	-----	386	QY	414	-----	413
Db	23667	KPESDGGSKI TGVVEMQTKGSEKSTCTQVKTLEATISGLTAGEBYVFRVAANEKGRS	23726	Db	24747	VCSI NSPSEVVGPI TCIDSYGGPVIDLPLEYTEVVKYRAGTSVKLRAGISGKPAPTI EWY	24806
QY	387	-----	386	QY	414	-----	413
Db	23727	DPQLGVPVIARDIEIKPSVELPFHTFNVKAREQLXIDVPFKGRPOATVNRKDGOTLKE	23786	Db	24807	KODKELQTNALVCVENTTDLASILIKADRLNSGCVELKLRNAMGSASATIRVQLDKPG	24866
QY	387	-----	386	QY	414	-----	413
Db	23787	TTRVNVSSKTVTSLSIKEASKEDVGTIELCVSNSAGSITVPIITIIIVDRPGPGPIRID	23846	Db	24867	PPGPIEFKTVTAEBKITLLWRPPADGGAKITHYIIVEKRETSRVVMSVSEHLEBECIIT	24926
QY	387	-----	386	QY	414	-----	413
Db	23847	EVSCDSITISNPPPYDGGCQISNYIIVEKETTSTTWHIVSQAVARTSIKIVRLTTGSEY	23906	Db	24927	TKLIIKNEVI FRVRAVNKYGICEPLESDSVAKNAFVTPGPPGIPVETKLTKNMTVVMS	24986
QY	387	-----	386	QY	414	-----	413
Db	23907	QFRVCAENRYKSSYSESSAVVAEYFPSPGPGTKVVAHATSTMLVTWQVPVNDGGR	23966	Db	24987	RPIADGGS DISGYFLEKRDKKSGLGWFKVLKETIRDTROKVTGLTENS DYQYRVCAVNAAG	25046
QY	387	-----	386	QY	414	-----	416
Db	23967	VIGYHLEYKERSILWSKANKILIA DTOMKVSGLDEGLMVEYRYVAENIAGICKSKSCE	24026	Db	25047	QGFSEPSSEFYKAADPIDPDPGPPAKIRIADSTKSSITLWNSKPVDGGS AVTGYVVEIRQ	25106
QY	387	-----	386	QY	417	-----	416
Db	24027	PVPARDCPPGQPEVTNITRKS VSLKWSKPHYDGGAKITGYIIVERREL PDGRWLKCNVT	24086	Db	25107	GEEB EWTTVSTKGEVRTTEYVVVSNL KPGVNY YFRVSAVNCAGQGBPIEMNEPVQAKDILE	25166
QY	387	-----	386	QY	417	-----	416
Db	24087	NIQETFEVTELTEQRYEPRVFARNAADSVSPSESTGPIIVKDDVEPRVNMVVKFRD	24146	Db	25167	APEIDL VALRTSIAKAGEDVQLIPFKGRPPPTVTRKDEKNLGS DARYSIENTDSSS	25226
QY	387	-----	386	QY	417	-----	419
Db	24147	VI VVKAGEVLKINADIAGRPLPVISWAKGIEIEERARTEIISTDNHTLLTVKDCIRRD	24206	Db	25227	LLTIPQVTRNDGKYILTIENGVBPKSSTVSVKVLDTPAACQKQVHKVSRGTVTLWD	25286
QY	387	-----	386	QY	420	-----	419
Db	24207	GQVYLT LKNVAGTRSAVAVNKVLDKPGPAGPLEINGLTAEBKCSLSWGRPOBGGADIDY	24266	Db	25287	PPLIDGSGPI INYVIEKRDATKRTWSVSHKCSSTSFKLIDLSEKTPPFRVLAENEIGI	25346
QY	387	-----	386	QY	420	-----	419
Db	24267	YIIVEKQPSPEKLGVTISISKDSVSLTLKPEHDGGSRIVHYVVVALEKQKNWVKAVA	24326	Db	25347	GBPCETTEPVKAAVPAPIRDLMSKDSTKTSVILSWTKDPDGGSVITEYVVERKGBGQ	25406
QY	387	-----	392	QY	420	-----	419
Db	24327	KSTHHVVSGLRENSEYFVRFAENQAGLSDPRELLLPVLIKEQLEPPEIDMKNFPSTVY	24386	Db	25407	TWSHAGISKTCEIEVSQLEKQSVLEFRVPAKNEKGLSDPVTIGPITVKELIITPEVDLSD	25466
QY	393	-----	392	QY	420	-----	419
Db	24387	VRAGSNLKVDIPISGKPLPKVTLSRDGVPLKATMRFNTEITAE NLINLKS VTA DAGRY	24446	Db	25467	IPGAQVTVRIGHNVHLELPYKGPKEPSISWLXDG LPLKESEFVRFSKTENKITLSIKNAK	25526
QY	393	-----	405	QY	420	-----	419
Db	24447	EITAANS GTTKAFINIVLDRPGPTGVPVWISDITEESVTLKWEPPKYDGGQVNYIL	24506	Db	25527	KEHGKGYTVILDNAVCRIAVPITVITLPGPSKPGPIRFDEIKADSVILSWDVPEDNGG	25586
QY	406	-----	405	QY	420	-----	419
Db	24507	LKRETSTAVWTEVSATVARTMMVKMLTTGGEYQFRIKAENRFGISDHIDSACVTVKLPY	24566	Db	25587	EITCYSIEKRETSQTNWKCVCSSVARTTPKPNLVKDAEYQFRVRAENRYGVSQPLVSSI	25646
QY	406	-----	405	QY	420	-----	419
Db	24567	TTPGPPSTPWNTVRESITVGHHEPVSGGS AVVGHLEMKDRNSILWOKANKL VIRT	24626	Db	25647	IVAKHQFRIPGPGKPEVIYNTVSDGMSLTWADAPVYDGGSEVTFGHVEKKERN SILWQKN	25706
QY	406	-----	SMG 408	QY	420	-----	419
Db	24627	HFKVTTISAGLIYEPRVYAENAGVKPSHPSEPVLAIDACEPPRNVRTIDISKNSVLS	24686	Db	25707	TSPISGREYRATCLVEGLDYQFRVYAENAGLS SPSDPKFTLANSPVDPGTPDYDVT	25766
QY	409	FQ-----	RTF-----	QY	420	-----	419
				Db	25767	RETITLKNPPLRDGGSKIVGYISIEKROGNERWRCNFTDVS E CQYTVTGLSPGDRYEPR	25826

QY	420	-----YDSLL-----	425
Db	25827	IIARNAVGTISPPSQSGIIIMTRDENVPPIVEFGPEYFDGLIITKSGESLRIKALVOGRPV	25886
QY	426	-----	425
Db	25887	PRVTWPKDGVIEIKRMMEITDVLGSTSLFVRDATDRHRGVTYVEAKNAGSASAKAEIKVK	25946
QY	426	-----	425
Db	25947	VQDTPGVVGPPIRFTNITGEKMTLWDAPLNDGCAPITHYIIEKRETSRLAWALIEDKCE	26006
QY	426	-----	425
Db	26007	AQSYTAIKLINGNEYQFRVAVNKGFGVGRPLDSPVVAQIQYTPDAPGIPESNITGNS	26066
QY	426	-----	425
Db	26067	ITLTWARPESDGGSEIQYILERREKXSTRWVKVISKRPISETRPKVTGLTEGNEYEFHV	26126
QY	426	-----	425
Db	26127	MAENAGVGSPASGISRLIKCREPWNPPGPTTVVKVTDTSKTTYSLEWSKPVFDGMEIIG	26186
QY	426	-----	425
Db	26187	YIEMCKADIADWHKVAEACVKTRYTVTDLQAGEYKFRVSAINGAGKGDSCVETGTIK	26246
QY	426	-----	425
Db	26247	AVDRLTAPELDIDANFKQTHVVRAGASIRLFIAYQGRPTPTAVWSKPDNSLSLRADIHTT	26306
QY	426	-----	425
Db	26307	DSFSTLTVCNCRNDAGKTYLTVENNNSGSKIITFTVKVLDTPGPGPITFKDVTGRSATL	26366
QY	426	-----	425
Db	26367	MWDAPLLDGGARIIHHYVVKREASRSWQVISEKTRQIPKVNDLAEGVPYPRVSAVNE	26426
QY	426	-----	425
Db	26427	YGVEBPEMPEPIVATEQPAPRRRLDVDTSKSSAVLMLKPDHDGGSRTGYLLEMRQK	26486
QY	426	-----	425
Db	26487	GSDFWVEAGHTKQLTFTVERLVEKTEYEPVXAKNDAGYSEPREAFSSVIIKEPQIEPTA	26546
QY	426	-----	425
Db	26547	DLTGITNQLITCKAGSPFTIDVPISGRPAPKVTWLEEMRLKETDRVSITTTKDRTLTV	26606
QY	426	-----	425
Db	26607	KOSMRDGSRYPLTLTENTAGVKTFSVTVVVIGRPGVPVGPBIEVSSVSAESCVLSWGEPKD	26666
QY	426	-----	425
Db	26667	GGGTEITNYIVKRESGTTAWQLVNSVVKRTQIKVTHLTKYMEYSPRVSSNRFVGSKPL	26726
QY	426	-----	425
Db	26727	ESAPIIAEHPFVPPAPTREVYVHSANAMSIRWEEPYHDGSKIIGYWEKKERTILW	26786
QY	426	-----	425
Db	26787	VKENVPCLCNKYVTGLVEGLEYPRTYALNAGVSKASEASRPIMAQNVPDAPGRPEV	26846
QY	426	-----	425
Db	26847	TDVTRSTVSLNSAPAYDGGSKVVGXIIBERKPVSEVGDGRWLKCNNTYIVSDNFTVTALS	26906
QY	426	-----	425
QY	426	-----	425
Db	26907	EGDTYFRVLAKNAAGVISKGSSESTGVTCDREYAPPAKELDARLHGLDVTIRAGSDLV	26966
QY	426	-----	425
Db	26967	DAAVGKPEPKIIWTGDKELDCEKVSLOYTKRATAVIKPCDRSDSGKYTLTVKNASG	27026
QY	426	-----	425
Db	27027	TKAVSMVKVLDSPGPGKLTVSRTVQEKCTLAWSLPQEDGGAEITHYIVERRETSRLNW	27086
QY	426	-----	425
Db	27087	VIVEGECPTLSVVVTRLIKNNEYIFRVRVKNYKGGVGVSESEPIVARNSTFIPSGPIPE	27146
QY	426	-----	425
Db	27147	EVGTGKEHIIIOWTKPESDGGNEISNYLVDKREKSLRWTRVNKDYYVVDYTRUKVTSLME	27206
QY	426	-----	425
Db	27207	GCDDYQFRVAVNAAGNSEPSEASNPISCREPSYTPGPPSAPRVVDTTKHSISLAWTKPMY	27266
QY	426	-----	425
Db	27267	DGCTDIVGYVLEMOEKDQDQWVRVHTNATIRNTEFTVPDLKMGQKYSFRVAANVKGMS	27326
QY	426	-----	425
Db	27327	YSESIAEIBPVERIBIPDLELADDLKKTVTIRAGASLRMLVSVSGRPPPVITWSKOGIDL	27386
QY	426	-----	425
Db	27387	ASRAIIDTTESYSLLIIVDKVNYDAGKYIEAENQSGKKSATVLVKVYDTPGPCPSVKVK	27446
QY	426	-----	425
Db	27447	EVSRSVTTIBIPTIDGAPVNNYIVKREAAWRAFKVTVTTCSTKTYRISGLVBGTM	27506
QY	426	-----	425
Db	27507	YFRVLPENIYGEBCEPCEPCEPCEPCEPCEPCEPCEPCEPCEPCEPCEPCEPCEPCEP	27566
QY	426	-----	425
Db	27567	GYVLEACKAGTERWKKVTLKPTVLEHTVTSLSNEGOYLFRIPAQNEKGVSEPREVTAV	27626
QY	426	-----	425
Db	27627	TVQDLRESERVTVETHKVAKLTIKRETTIRDTGTYLTKNVTGTTSETIKVILDKPGP	27686
QY	426	-----	425
Db	27687	PTGPIKIDEIDATISISWEPPELPGAPLSGVVVEQDRAHRLPGLVPSVESVTRSTFKFT	27746
QY	426	-----	425
Db	27747	RLTEGNEYFRVAATNRFIGISYLOSEVIECRSSIRIPGPPEPQLQIFDVSRDGMTLTWYP	27806
QY	426	-----	425
Db	27807	PEDDGSQVQTYIVERKEVRADRWVRVNVKVPVTMTYRSTRGLTEGLEYEHRVTAINARGS	27866
QY	426	-----	425
Db	27867	GKPSRPSKPIVAMDPIAPPGKPNPRVTDTRTSVLSAWSVPEDEGSKVTGYLIBQKV	27926
QY	426	-----	425
Db	27927	DOHEWTKCNTTPTKIREYTLTHLPOGAEYFRVLACNAGGPGEPBVPCTVKVTEMLEYP	27986
QY	426	-----	425

Db	27987	DYELDERYQEGIFVRQGGVIRLTPIPIKGPFPICKWTKEGQDISKAMIATSEHTELVI	28046	Db	29067	KEDKTRAMNYDEEVDRETVSMTKASHSSTKELYKYMIAEDLGRGEFGIVHRCVETSSK	29126
Qy	426	-----	425	Qy	437	-----	436
Db	28047	KEADRGDSGTDLVLEKNCKGKXAVYIKRVIGSPNSPEGPLEYDDIQRVSRVSWRPPAD	28106	Db	29127	KTYMAKFVKVGTDOVLVKKEISILNIARHRNHLHSHESFESMEELVMIFESIGLDIFE	29186
Qy	426	-----	425	Qy	437	-----	436
Db	28107	DGGADILGYILERREVPKAAWYITIDSRVGTSLVVKGLKENVEYHFRVSAENQFGISKPL	28166	Db	29187	RINTSAFELNEREIVSVYHVQCEALQFLSHNHIGHDIRPENIYYOTRRSSTIKIIEFGQ	29246
Qy	426	-----	425	Qy	437	-----	436
Db	28167	KSEEPVTKPLNPPEPPNPPELVDTKSSVLSWSRPKDDGGSVRTGYIIERKETSTD	28226	Db	29247	ARQLKPGDNFRLLFTAPEYAPAEVHQHDVVSTATDMWSLGTLYVLLSGINPFLAETNOQ	29306
Qy	426	-----	425	Qy	437	-----	436
Db	28227	KWVRHNKTQITTTMYTWTGLVPDABYQFRIIAQNVDGLSETSPASEPVPVCKDPFDKPSQP	28286	Db	29307	IINENMAEYTFDBEAFKEISIEAMD FVDRLLVKERKSRMTASEALQHPWLKQKIERVST	29366
Qy	426	-----	425	Qy	437	-----	436
Db	28287	GELEILSISKDSVTLOWKPECDGGKEILGYWVEYRQSGDSAWKKSNERIKDKQFTIGG	28346	Db	29367	KVIRTLKHRRYYHTLIKKDLNMVVSAAIRSCGAIIRQSVSVAKVVASIEBIGVSGOI	29426
Qy	426	-----	425	Qy	437	-----	436
Db	28347	LLEATEYFRVFAENETGLSRPRTAMSIKTKLTSGEAPGIRKEMKDVTTKLGEEAQLSC	28406	Db	29427	MHAVGEGGHVKYVCKIENYDOSTQVTVFQVRQLENSEKEYEITYEDGVAIIYVKDITKL	29486
Qy	426	-----	425	Qy	437	-----	436
Db	28407	QIVGRPLDPDKWYRFKELIQSRKYKMSDSGRHTLTVMTEBOEDSGVYTCIATNEVGEV	28466	Db	29487	DDGTYRCKVNDYGEDSSVAELFVKGVREVYDYCRRTWKIKRRTDTMRLLERPPEFTL	29546
Qy	426	-----	425	Qy	437	-----	436
Db	28467	ETSSKLLQLQATPOFHPGYPLEKXYGAVGSTLRHLWYVIGRPVPMATWFGQKLLQNSN	28526	Db	29547	PLYNKTA YGVGNVRFVTITVHPPEHVTWYKSGQIKPGDNDKKYTFESDKGLYQLTINS	29606
Qy	426	-----	425	Qy	437	-----	436
Db	28527	ITIENTEHYTHLVKMNQVRKTHAGKYQVLSNFGTVDAILDVEIQDKPKPTGPIVIEA	28586	Db	29607	VTTDDAEYTVVARNKYGEDSKAKLTVTLHPPPTDSTLRPMFKRLLANAECEQGSVCF	29666
Qy	426	-----	425	Qy	437	-----	436
Db	28587	LLKNSAVISWKPADDGGSWITNYVEKCEKEGAEWQLVSSAISVTTCRIVNLTENAGY	28646	Db	29667	EIRVSGIPPTLKWEKGQPLSLGPNIEIIHEGLDYVALHRLDTPEDTGYVRVTATNTA	29726
Qy	426	-----	425	Qy	437	-----	436
Db	28647	YFRVSAQNTFGISDPLEVSSWIIKSPFEKPGAPKPTITAVTKDSCVVAWKPPASDGA	28706	Db	29727	GSTSCQAHQVBLRYKKQEFKSEHEHRVOKIDKTLRMAEILSGTESVPLTQVAKEA	29786
Qy	426	-----	425	Qy	437	-----	436
Db	28707	KIRNYLEKREKQKQWISVTTBEIRETVFSVKNLIEGLEVEYFRVKCNLGGSEWSEIS	28766	Db	29787	LREAAVLKPAVSTKTVKGEFPLEIEEKKERKLMPYDVPEPRKYKQTTIEEDQRIKQF	29846
Qy	426	-----FLDEIRNFSLR-----	436	Qy	437	-----	436
Db	28767	EPITPKSDVPIQAPHFKBELNLNRYQSNATLVCKVTGHPKPIVKWYRQGEKIIADGLK	28826	Db	29847	VEWSDMKWYKIRDOQVEMPGKLDREVVKRPRKIRLSRWELLYQVRPKWRIPKLSQDDLE	29906
Qy	437	-----	436	Qy	437	-----	436
Db	28827	YRIQEFKGGYHQLIIASVTDDATVYQVRATNOGQSVSGTASLEVEVPKIHLPKLTLEGM	28886	Db	29907	IVRPARRTPSPDYFYVRPRRSLGDISDELLLPIDDDYLAWKRTTEERLRLEEBELG	29966
Qy	437	-----	436	Qy	437	-----	436
Db	28887	GAVALRGEVVISIKIPSGKPDVITWQGGDLIDNNGHYQVIVTRSFVSLVPNGVERK	28946	Db	29967	FSASPPSRPPHPELSSLYSSPQAHVVEETRKDFRYSTYHIPTKAEASTSYAELRERH	30026
Qy	437	-----	436	Qy	437	-----	436
Db	28947	DAGFYVVCANRRFGIDQKTVELDVADVPDPPRGVKVSDVSRDSVNLTWTEPASDGGSKIT	29006	Db	30027	AQAAYRQPKQRQIRMAEREDELLRPVTTTQHLSEYKSELDFMSKEEKSRRQREVT	30086
Qy	437	-----	436	Qy	437	-----	443
Db	29007	NYIVEKATTAEWLVRVGQARETRYTVINLFGKTSYQFRVIAENKFGLSKPSBPSEPTIT	29066	Db	30087	EITEIEEVEISKHAQRESSSASRLRRRSLSPITYIELMRPVSELIRSRQPABEYED	30146
Qy	437	-----	436	Qy	444	-----TP-----	445
				Qy	30147	DTERRSPTERTPRSPSPVSSERSISRFERSARPDIFSRYESMKAALKTQKTSERKYEY	30206

```
QY 446 ----- 445
Db 30207 LSQPFTHDAPRITILMRSHRVPCCGQNTFRILNVQSKPTAEVKWYHNGVQLQESSKIHY 30266
QY 446 ----- 445
Db 30267 TWTSGVLTLEILDCHTDDSGTYRAVCTNYKGEASDYATLDVTGGDYTTVASQRRDEVRP 30326
QY 446 ----- 445
Db 30327 SVPPELTRTEAVSFKTSEMEASSVREKVSQMTETRESLSSYEHSAEMKSAALE 30386
QY 446 ----- 445
Db 30387 EKSLEBKSTTRKIKTTLAARILTKPRSMVTYEGESARFSCDTGDEPVTWLRKGQVLS 30446
QY 446 ----- 445
Db 30447 TSARHQVTTTKYSTPEISSVQASDEGNYSVVENSEGKQAEFTLTIOKARVTEKAVTS 30506
QY 446 ----- 452
Db 30507 PPRVKSPEPRVKSPEAVKSPKRVKSPSPKAVSPTEKPTPEKQVHLVSPAPKIQ 30566
QY 453 ----- NLS 455
Db 30567 FLKAEASKIEIAKLTGVSSVLRKEVTWYDQKCLKENGHFQFYSADGTVELKINNLT 30626
QY 456 TSN----- 458
Db 30627 ESDQGEYVCEISGEGGTSKTNLQFMGQAFKSIHEKYSKISSETKSKDQKTTSTVTRKTP 30686
QY 459 ----- SLW----- 461
Db 30687 KAPEPISSKPVIVTGLQDTTVSSDSVAKPAVKATGEPRPTAIWTKDKAITOGGKYLSE 30746
QY 462 ----- 461
Db 30747 DRGGFLEIHKTDSDGLYCTVKNAGSVSSCKLTIKAIKDTBAQVSKTKTSEITP 30806
QY 462 ----- 461
Db 30807 QKXAVVQEEISOKALRSEIKMSEAKSQBKALKEEASKVLISEEVKKSAAATSLKSIYH 30866
QY 462 ----- 462
Db 30867 EBITKSQASEEVRTHAEIKAFSTQMSINEGQRLVLKANIAGATDVKKVNLGVELTNSEE 30926
QY 463 ----- 462
Db 30927 YRYGSGSDQTLTIKQASHRDEGILTCISKTEGIVKCOYDLTSLKSLSDAPAFISQPRS 30986
QY 463 ----- WLQ 465
Db 30987 QNINEQNVLFTCEISGEFSPRIEWFK 31013
```

RESULT 4
ADQ17316
ID ADQ17316 standard; protein; 26926 AA.
XX AC
XX ADQ17316;
XX AC
DT 26-AUG-2004 (first entry)
XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 133.
XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX OS Homo sapiens.
XX PN W02004048938-A2.

```
XX 10-JUN-2004.
PD 26-NOV-2003; 2003WO-US038193.
XX 26-NOV-2002; 2002US-0429739P.
PR (PROT-) PROTEIN DESIGN LABS INC.
XX Aziz N, Ginsburg WM, Zlotnik A;
PI WPI; 2004-441208/41.
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 133; 210pp; English.
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 26926 AA;
Query Match 27.8%; Score 681; DB 8; Length 26926;
Best Local Similarity 1.0%; Pred. No. 1.2e-06;
Matches 268; Conservative 91; Mismatches 106; Indels 25559; Gaps 84;
QY 1 MGRK----- 4
Db 649 VGKKAQAVTVAVDQARVREPFGHLEBSYAQQTLEYGYKERISAQVAPQPPA 708
QY 5 ----- 4
Db 709 SEPHVVPKAVPRVIAQSETHIKTTDQGMHISQIKKTTDLTTERLVHVDKRPRTASP 768
QY 5 ----- 4
Db 769 HFTVSKISVPKTEHGYEASIAISATLQKELSATSSAOKITKSVKAPTVPKPTRVRAE 828
QY 5 ----- 4
Db 829 PTPLPQFPFADPDYKSEAGVEVKVEGVSITGTTVREERFEVLHGREAKVTETARVPA 888
QY 5 ----- 4
Db 889 PVEIPVTPPTLVSLGNVTIVIEGESVTLECHISGYPSPTVTVREDYQIESIDFQITQ 948
QY 5 ----EMVVRD----- 10
Db 949 SGIARLMIREAPAFEDSGRFTCSAVNEAGTVSTCYLAVQVSEFEKETTAATEKFTTEEK 1008
QY 11 ----- 10
Db 1009 RFVESRDVVMVMTSLTEBQAGPEPAAPYFITKPVVQKLVEGGSVFGQVGNPKPHYV 1068
QY 11 ----- 10
Db 1069 WKXSGVPLTTGYRYKVYNKQGTGECKLVISMTFADDAGEYTIWVKNKHGETSASASLLEE 1128
QY 11 ----- 10
```

Db 1129 ADYELLMSQQEWLYQTVTAQVQPEVGETAGPVFVSEYKEYEKEQALIRKQWAKDTV 1188
QY 11 ----- 10
Db 1189 VVRTVVEOEFHISSEERLIKEIVRIIKITTELLBEDGEBKMAVDISESEAVESGFD 1248
QY 11 -----VPK----- 13
Db 1249 LRIKNVRILEGWVTHCKWGWYLPFKIAWYKDGKRIKHGERYQMDFLQDGRASLRIPV 1308
QY 14 ----- 13
Db 1309 LPDEGIYTAFAASNIKGNACSGKLYVEPAAPLGAFTYIPTLEPVSRIRSLSPRSVSRSP 1368
QY 14 -----MFVLISISFLVSPNC 30
Db 1369 IRMSPARMSPARMSPARMSPOGRLEETDESOLRLYKPVFLKPVSPKLEGANC 1428
QY 31 -----KWSK----- 35
Db 1429 RFDLKVVGRMPETFWHDDQQIVNDYTHKWKIKEDGTQSLIIVPATPSDSEWTVVAQN 1488
QY 36 ----- 35
Db 1489 RAGRSSIVILTVEAVEHQVPMFVEKLNVINKEGSRLEMKVRAKGNPNPDIVLKNSD 1548
QY 36 ----- 35
Db 1549 IIVPHVKPIRIBGTGEAALKIDSTVQSDSAWYTATATINKAGRDTTRCKVNVVEFAP 1608
QY 36 ----- 35
Db 1609 EPERKLIIPRGYTRAXEIAAPELEPLHLRYQGEWEGDLYDKEKQKPPFKKLTSLRL 1668
QY 36 ----- 35
Db 1669 KRFGPAHFECRLTPI SDPTMVVWELHDGKPLEAANRLMINEFGYCSLDYGVAYSRDSGI 1728
QY 36 ----- 35
Db 1729 ITCRATNKYGTDHTSATLIVKDEKSLVESQLEGRKGLQRIEELERMAHEGALTGVTTD 1788
QY 36 -----ALXNRPR----- 43
Db 1789 QKEKQKPDIVLYPEFVRVLEGETARFCRVGYQPKVWYNGQLIRKSRFRVYDGI 1848
QY 44 -----GLVLSKI----- 50
Db 1849 HYLDIVDCKSYDTGEVKVTAENPEGVIEHKVKLEIQOEDFRSVLRRAPRPEFHVHEP 1908
QY 51 ----- 50
Db 1909 GKLOFEVQKVDPRPVDTTTETKEVVKLKRAIRITHKVPSESEELRSKFRRTESGYEAIT 1968
QY 51 ----- 50
Db 1969 AVELSKRKQDSYEELLARKTKDELLHWTKELTBEKKALAEKGKITIPTFKPKDIELSPS 2028
QY 51 ----- 50
Db 2029 MEAPKIPERIQSTVQGSDAHFRVRVVGKPDPECEWYKNGVKIERSDRYIYWMPEDNVC 2088
QY 51 ----- 50
Db 2089 ELVIRDVTAEDSASIMVKAINIAGETSSHAFLLVQAKQLITFTQELQDVVAKEDWTATF 2148
QY 51 ----- 50
Db 2149 ECETSEPFVKVYKDGMEVHEGDKYRMHSDRKVHFLSILITDSDAEDYSCVLVEDENV 2208
QY 51 -----GKYKL 55
||| :

Db 2209 KTTAKLIVEGAVVEFVKELQDIEVPSSYSGELECIIVSPENIEGKWHNDVELKSNKYTI 2268
QY 56 -----DQLK 59
Db 2269 TSSRRQNLTVKDVTKEDQGEYSFVIDGKTTCKLKMKPRPIALQGLSQKQCEGDIVQ 2328
QY 60 LEILRQLET-----TI----- 70
Db 2329 LEVKVLSVESGVMMKQGEVQPSDRVHIVIDKQSHMLLIETKEDAGNYSFTIPALG 2388
QY 71 ----- 70
Db 2389 STSGRVSVSYVDVITPLKDVNVNIEGTAKVLECKVSPDVTSVKMYLNDEQIKPDDRVQAI 2448
QY 71 ----- 70
Db 2449 VKGTQORLVINRTHASDEGPKLIVGRVETNCNLSVEKIKIRGLRDLTCTETQNVVFEV 2508
QY 71 -----STKYNVS----- 77
Db 2509 ELSHSGIDVLNPKDKBIKPSKYKIEAHGKIYKLTVLNMMKODEGKYTFYAGENMTSGK 2568
QY 78 ----- 77
Db 2569 LTVAGGAIKPLTDQTVAESQEAUFCEVAVPDSKGEWLRDGHPLPLTNNIRSESDGHR 2628
QY 78 -----KOPVKNLTM----- 86
Db 2629 RLIIAATKLDIDIGVTVKATSKTSAKLKVEAVKIKTKLNLITVTETQDAVTVVELTHPN 2688
QY 87 ----- 86
Db 2689 VKGVQWIKNGVLSNEKYAISVKGTIYSLRIKCAIVDESUYGFRGLRGLSARLHVET 2748
QY 87 ----- 86
Db 2749 VKIIKPKDVTALENATVAFESVSHDTPVKWFHKSVEIKPSDKHRLVSRKVKHKLMLQ 2808
QY 87 -----NTEFPQ----- 92
Db 2809 NISPSDAGEYTAVVQGLECKAKLFVETLHITKTMKNIIEVPETKTASPECEVSHFNVPMSW 2868
QY 93 ----- 92
Db 2869 LKNGVEIEMSEKFKIVQGLHQLIIMTSTEDSABYFVCGNDOVSATLITVPIMITSM 2928
QY 93 ----- 92
Db 2929 LKDINAEKDTITPEVTNVEGISYKWLKNGVEIKSTDKCQMRTKKLTSLNIRNVHFGD 2988
QY 93 ----- 92
Db 2989 AADYTFVAGKATSTATLYEARHIEFRKHIKD IKVLEKKRAMFECEVSEPDITVQMMKDD 3048
QY 93 -----YYILAG----- 98
Db 3049 QELQITDRIKIOKEKYVHLLIPSTRMSDAGKYTVVAGNVSTAKLFVEGRDVRIRSICK 3108
QY 99 ----- 98
Db 3109 EVQVIEKQRAVVEFEVNEDDVDHAWYKDGIIEINFQVBRHKYVVERIRHMFISETRQSD 3168
QY 99 ----- 98
Db 3169 AGEYTFVAGNRSSVTLYVNAPEPPVLOELQPVTVQSGKPARFCAMISGRPOPKLSWYK 3228
QY 99 ----- 98
Db 3229 BEQLSTGPKCKELHDGQBYTLLIEAFPEDAAYVTCBAKNDYGVATTASLSVEVPEV 3288
QY 99 -----PIQN----- 102
Db 3289 SPDQEMPVYPPAIITPLQDVTVSEGPARFCQVRSGTDLKVSWSKDKKIKPSRFRMTQ 3348
||| :

QY 103 ----- 102
Db 3349 FEDTYQLEIAEAYPEDEGTYTFVANNAGVQSVSTANLSLEAPESILHERIEQIEMEMKE 3408
QY 103 ----- 102
Db 3409 FSSSFLSABEEGHSALQSKINETLELLESFVPTKPDSEKGTGPIFIKEVSNADI 3468
QY 103 ----- 102
Db 3469 SMGDVATLSVTVIGIPKPKIQWFNGVLLTPSADYKFEVPGDDHSLIILFTKLEDEGEYT 3528
QY 103 ----- 102
Db 3529 CMASNDYGKTCISAYLKINSKGBGHKDTETESAVAKSLEKLGCPHFKELKPIRCAQ 3588
QY 103 ----- 110
Db 3589 GLPAIFEYTVVGPAPTFTWFKENKQLCTSVYVYTIHNPNGSGTFIVDPQREDSGLYIC 3648
QY 111 ----- 121
Db 3649 KAENMLGESTCAEALLVLEDDTMDTTPCKAKSTPEAPDFPQTPLKGPFAVEALDSEQEI 3708
QY 122 ----- 121
Db 3709 ATPVKDTILKAALITEENQQLSYEHIAKANELSSQLPLGAQELQSILOQDKLTPESTREF 3768
QY 122 ----- 121
Db 3769 LCINGSIHFOPLKEPSNQLQIVSQKTSKEGILMPPEPQAVLSDEKIFPSAMSI 3828
QY 122 ----- 121
Db 3829 EQINSLTVPLKTLAEPGNYPOSSIEPPMHSYLTSVAEVLSLKEKTVSDTNREQRTV 3888
QY 122 ----- 121
Db 3889 LQKQEAQSALILSQSLAEGHVESLQSPDVMISQVNYEPLVPSHSCTEGGKILIESANPL 3948
QY 122 ----- 121
Db 3949 ENAGQSAVRIERGKSLRFPPLALEEKQVLLKEHSDNVVMPDQIIIESKREPVAIRKQOE 4008
QY 122 ----- 121
Db 4009 VQGRDLLSKESLLSGIPBEORLNLKIQICRALQAAVASQPGFLFSEWLNRNIEKVEAVN 4068
QY 122 ----- 121
Db 4069 ITQEPRHNCMYLVSATKSVEEVTIIIEDVDPMANLKMELRDALCAIIEIDILTAE 4128
QY 122 ----- 121
Db 4129 GPRIQQAKTSLOEEMDSFSGQKVPITEPEVESKYLISTEVSFYFNQSRVKYLDATP 4188
QY 122 ----- 121
Db 4189 VTKGVASAVVSDKQDESUKPSEKESSSESCTEETVATVKIQEAGGLIKEDGPMIHTP 4248
QY 122 ----- 121
Db 4249 LVDTVSEEGDIVHLTTSITNAKEVNYFENKLVPSDEKFKCLQDQNTYTLVIDKVNTEDH 4308
QY 122 --KYVYSQVNHAKTI-
: : : : :
Db 4309 QGEYVCEALNDSKGTATSAKLTVVKGAAAPVIKRIBPLEVALGHLAKFTCEIQSAPNRF 4368
QY 136 ----- 135
Db 4369 QWPKAGREIYESDKSIRSKYITSSIEILRTQVDDCEYTCASBYSGVSTATLTVTV 4428

QY 136 ----- 135
Db 4429 PGSEKKVRKLLPERKPEPEKEEVLKSVLRKPEEBEPKVEPKLEKVKYKPAVPEPPPKP 4488
QY 136 ----- 141
Db 4489 VEEVEVPTTKBERKIPEPTKVPEIKPAIPLAPBPKPKPEAEVTKIKPPVPPEPTPIA 4548
QY 142 ----- 141
Db 4549 APVTVPVVGKAEAKAPKEEAAKPGPIKGVPKTSPIEABRRKLRPGSGGKPPDEAP 4608
QY 142 ----- 141
Db 4609 FTYQLKAVPLKFVKEIKOIIILTESEFVGSSAIFECIVSPSTAITTMMKGSNIRESPKHR 4668
QY 142 ----- 141
Db 4669 FIADGKDRKHLIIDVQLSDAGEYTCVLRGNKEKSTAKLVVEELPVRPVKTLSEETVV 4728
QY 142 ----- 145
Db 4729 KGOPYLSCELNKERDVWRKDKIIVVEKPGRIVPGVIGLMRALTINDADDDTAGTYTVT 4788
QY 146 ----- 145
Db 4789 VENANNLECCSCVKVVEVIRDMVKPIRDQHVKPKGTAFACDIAKDTPNIKWFKGYDEI 4848
QY 146 ----- 145
Db 4849 PAEPNDKTEILRDGNHLYDKIRNAMPEDIAEVAVEIEGKRYPAKTLGSEVELLAKPIED 4908
QY 146 ----- 150
Db 4909 VTIIYKESASFDAEISEADI PGQWKLKCELLRSPSTCEIKAEKGKRFLLHKVKLDQAGE 4968
QY 151 ----- 150
Db 4969 VLYQALNAITTAITLVKEIELDPVPLKDVTPERRQARFECVLTREANVINSGPDIK 5028
QY 151 ----- 150
Db 5029 SSDKFDIIADGKKHILVINDSQFDDGEVYTAEVGKTSARLFTVGTIRLKFMSPLSDQTV 5088
QY 151 ----- 150
Db 5089 KEGETATFVCELSHSHQVHVWFKNDAKLHSTRTVLISSEKTHKLEKVTLLDDISQIKA 5148
QY 151 ----- 150
Db 5149 QVKELSSSTAQLKVLBADPYFTVKLHDKTAVEKDEITLKCEVSKDVPKWKDGEIIVPSP 5208
QY 151 ----- 150
Db 5209 KYSIKADGLRRLIKTKADLKDGEYVDCGTDKTKANVTVEARLIEVEKPLYGVFVVG 5268
QY 151 ----- 150
Db 5269 ETAHPEIELSEPDVHGQWKLKGOPLTASPDCEIIEDGKKHILLHNCQLGTMTEVSFQAA 5328
QY 151 ----- 150
Db 5329 NAKSAANLKVKBPLIFITPLSDVKVFEKDEAKFECEVSREPKTFRWLKGTQBITGDDRF 5388
QY 151 ----- 150
Db 5389 ELIKDGTKHSVMIKSAAFEDEAKYFMEADKHTSGKLIIEGIRLKLFTPLKDVTAKEKS 5448
QY 151 ----- 150
Db 5449 AVPTVELSHDNIRVKWFKNDQRLHTRTSVSMQDEGKTHSITFKDLSIDDTSQIRVEAMGM 5508
QY 151 ----- 150

Db	5509	SSEAKLTVLEGGPYFTGKLQDYGVEKDEVILQCEISKADAPVKWFKDGKEIKPSKNAPI	5568	Db	6589	VIDVPGVRNLEVTETFDGEVSLAWEBPLTDGSGKIIGVYVERRDIKRTWVLTADRAES	6648
Qy	151	-----	150	Qy	155	-----LNY-----	157
Db	5569	KTDGKRMILILKALKSDIGQYTCGGTDTSGKLDIEDREIKLVRPLHSVEVMEETAR	5628	Db	6649	CEFTVTGLQKGVEYLFVRSARNRVCTGEPVETDNPVEARSKYDVPGLNVITITDVNR	6708
Qy	151	-----	150	Qy	158	-----	157
Db	5629	PETEISEDIDHANWKLKGEALLQTPDCEIKBEGKIHSVLHNCRLDQDGGVFOAANVKS	5688	Db	6709	GVSLTWBPEYDGCABITNVVIELRDKTSIRWDTAMTVRAEDLSATVTVVGEQESFRV	6768
Qy	151	-----	150	Qy	158	-----	157
Db	5689	SAHLRVKPRVIGLLRPLKDVTVTAGETATFDCELSYEDIPVEWYLGKKLBPSDKVPRS	5748	Db	6769	RAQNRIGVGKPSAATPFVKVADPIERPSPVNLTSDDQTSQSSVQLKWEPLKDGSGPILG	6828
Qy	151	-----	150	Qy	158	-----	157
Db	5749	EGKVHILTLRDVKLEDAQEVQLTAKDFKTHANLFVKEPPVEFTKPLEDQTVBEGATAVLE	5808	Db	6829	YIIERCEEGKDNWIRCNMKLVPELTVKYTGLEKGNKYLYRVSABNAGVSDPSEILGLPT	6888
Qy	151	-----	150	Qy	158	-----	157
Db	5809	CEVSRENKVKPFNGTEILSKKYEIVADGRVRLVIHDCPTEDIKTYTCDAKFKTSC	5868	Db	6889	ADDAFVBTMDLSAFKDGLEVIVPNPITILVPSTGYPRPTATWCFGDKVLETCGRVKMKT	6948
Qy	151	-----	150	Qy	158	-----	157
Db	5869	NLNVVPHVFLRPLTDLQVREKEMARPCBELSRENKVKWFKOGAEIKKKKYDIISKG	5928	Db	6949	LSAYAEELVISPERSDKGIYTLKLENRVKTSISGEIDVNVIAFPSAPKELKFGDITKDSVH	7008
Qy	151	-----	150	Qy	158	-----	157
Db	5929	AVRILVINKLLDDEABEYSCVTRTARTSGMLTVLEBEAVFTKNLANIEVSETDTIKLVCE	5988	Db	7009	LTWBPPDDGSGPLTGYVVEKREVSRTKTVKMDFVTDLEFTVPDLVQKKEYLKFVCARN	7068
Qy	151	-----	150	Qy	158	-----	157
Db	5989	VSKPGAELVWYKGBEIIETGRYEILTEGRKRILVITQNAHLEAGNVCNCLPSRDTGKY	6048	Db	7069	KCGEPAYVDEPNMSTPATVPPPENVKWRDTANSIFLTWDPKNDGSGRIKGYVE	7128
Qy	151	-----	150	Qy	158	-----	157
Db	6049	KVHELAAEFISKPNLEILEGEKAEFVCSISKESFPVQWKRDDKTLESQDKYDVIADGKK	6108	Db	7129	RCPRGSDKWWACBPVAAETKMEVTGLEEGKWYAYRVKTLNRQGASKPSRPTTEEIOAVDTQ	7188
Qy	151	-----	150	Qy	158	-----	157
Db	6109	RVLVVKQATLQDMGYVMVGAAARAAHLTVIEKLRIVVPLKOTRVKEQEVVFNCEVNT	6168	Db	7189	EAPFIFLDVKLAGLTVKAGTKIELPATVTKGPEPKITWTWKADMLLKQDKRITTIENVPKK	7248
Qy	151	-----	150	Qy	158	-----SKRNDTG-----	164
Db	6169	EGAKAKWFRNEAIFDSSKYIILQKDLVYTLIRDAHLDDQANYVSLTNHNGENVKSAA	6228	Db	7249	SIVTIVDSKRSDTGYTIIIEAVNVCGRATAVVEVNVLDKPGPPAAFDITDVTNESCLLTWN	7308
Qy	151	-----	150	Qy	165	-----	164
Db	6229	NLIVBEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLNVLTKWTKNGEVPFNDNRVSYRVDK	6288	Db	7309	PPRDDGSGKITNYVVERRATDSEVWHKLSSTVKOTNFKATKLIIPNKEYIFRVAENMYGA	7368
Qy	151	-----	150	Qy	165	-----	164
Db	6289	YKHMLTIKDCGPPDRGEVIVTAGQKSVAEALLIIEAPTEFVEHLEDQTVTEFDADVSCQ	6348	Db	7369	GEFVQASPIKACYQDPFGPPTRLBPSDITKDAVLTWCEPDDGSGSPITGYVVERLDPD	7428
Qy	151	-----	150	Qy	165	-----	164
Db	6349	LSREKANVWYRNGREIKEGKYPKEKDGSIHRLIHKDCLDDECEYACGVEDRKSARL	6408	Db	7429	TDKWVRCKMPVKDITYRVKGLTNKKYRFRVLAENLAGPGKPSKSTBILIKDPIDPPW	7488
Qy	151	-----LSEM-----	154	Qy	165	-----	164
Db	6409	FVEEIPVEIIRPPQDILEAPGADVFLAELNKDKVEVWLNNVVVQGDQKQMSBGI	6468	Db	7489	PPGKFTVKDVGKTSVRLNWKPEHDDGAKIESYVIEMLKTGTDEWVRVAEGVPTTQHLLP	7548
Qy	155	-----	154	Qy	165	-----	164
Db	6469	HRLOICDIKPROQGEYRPIAKDEARAKLELAAAPKIKTADQDLVVDVGKPLTMVVPYDA	6528	Db	7549	GLMEGQESYFRVAVNKAEGSEBSPDPVLCREKLYPPSPRWLEVINITKNTADLKWT	7608
Qy	155	-----	154	Qy	165	-----	164
Db	6529	YPKAEAWFKENEPSTKTIIDTAEQTSFRILEAKKGDGRYKIVLQNHKGKAEGFNLK	6588	Db	7609	VPEKDGSGSPITNYIVBKRDRVRKGWQTVDTTVKDKTCTVPLTEGSLYFRVAENAIQ	7668
Qy	155	-----	154	Qy	165	-----	164
Db	7728	SDYTEIEDSVLAKDTFTTGGPPYALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKSR	7728	Db	7669	SDYTEIEDSVLAKDTFTTGGPPYALAVDVTKRHVDLKWEPKNDGGRPIQRYVIEKKSR	7728

QY 165 -----BOGCG----- 169
Db 7729 LGTRWVAGTAGDCNFRVTDVIEGTEVQVRAENEAGVGHPSBTEILSIEDTSP 7788
QY 170 ----- 169
Db 7789 SPPLDLHVTDAGRKHIAIAWKPEKGGSPFIIGHVEMCPVGTEKWRVNSRPIKOLKFK 7848
QY 170 ----- 169
Db 7849 VEEGVDPKEYVLRVRAVNAIGVSEBSEISENVVAKDPDCKPTIDLETHDIIIVIEGEKLS 7908
QY 170 ----- 169
Db 7909 IPVFFRAVPVPTVSHKHGKEVKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKL 7968
QY 170 ----- 169
Db 7969 GSATASINVKIGLPGCPCKDIKASDITKSSCKLTWPEPPDGGTPTILHYVLERREAGRRT 8028
QY 170 ----- 169
Db 8029 YIPVMSGENKLSWTVKDLIPNGEYFFRVKAVNKGGEYIELKNPVAQDPKQPPDPVD 8088
QY 170 ----- 169
Db 8089 VEVHNPTEAMTITWKPLYDGGKIMGYIIIEKIAKEERWKRNEHLVPILTYTAKGLE 8148
QY 170 ----- 169
Db 8149 EGKEYQFRVRAENAAGISEPSRATPTTKAVDIPADPKVILRTSLEVRKGDEIALDASISG 8208
QY 170 ----- 169
Db 8209 SPYPTITWIKDENIVPEEIKKRAAPLVRRRKGEVEEPPVLPLOTLSIDNSKKGESQ 8268
QY 170 ----- 169
Db 8269 LRVRDSLRLPDHGLYMIKVENDHGIKAPCTVSLDTPGPPINFVEDIRKTSVLCKWEPP 8328
QY 170 -----NFT----- 172
Db 8329 LDGGSSEIINYLTLEKKDKTKPDSEWIVVSTLHRCKYSVTKLIEGKEYLFRVRAENRFGP 8388
QY 173 -----TFNPMF----- 178
Db 8389 GPPCVSKPLVAKDPFGPPDAPDKPIVEDVTSNMLVKNWEPKONGSPILGYWLEKEVNS 8448
QY 179 ----- 178
Db 8449 THMSRVNKSLLNALKANVDLGLLEGLTVFRVCAENAAAGPKFSPSPDPKTAHDPISPPGP 8508
QY 179 ----- 178
Db 8509 PIPRVTDTSSTTIELEWEPFAPNGGEIVGYFVDKOLVGTNKSCTERKMIKRVQYTVKE 8568
QY 179 ----- 178
Db 8569 IREGADYKLRVSAVNAAGEPPGETQPTVAEPQEPFAVELDVSVKGGIOMAGKTLRIP 8628
QY 179 ----- 178
Db 8629 AVVTGRPVTKWTKEEGELDKDRVIDNVGTKSELIIKDALKRDKHCRYVITATNSCGSK 8688
QY 179 ----- 178
Db 8689 FAAAARVEFVDPGVPLDKPVTNRKMLNWSDEPDGGEITGFIERKDAKMTWRQ 8748
QY 179 ----- 178
Db 8749 PIETERSKCDITGLLEGEQYKFRVIAKNKFGCGPPVEIGPILAVDPLGPPTSERLYTE 8808

QY 179 ----- 178
Db 8809 RORSTITLDWKEPRNSGGSPIOGYIIIEKRHRDKDPFERNKRLCPTTSPFLVNLDEHOMY 8868
QY 179 ----- 178
Db 8869 EFRVKAINEIGSESPSLPLNVVQDDEVPTIKLRLSVRGDTIKVKAQEPVHIPADVTGL 8928
QY 179 ----- 178
Db 8929 PMPKIEMSKNETVIEKPTDALQITKEEVSERSEAKTELSIPKAVREDKGTVTVTASNRLGS 8988
QY 179 ----- 178
Db 8989 VFRNVHVEVYDRPSPRNLAVTDIKAESCYLTDAPLDNGSGSEITHYVIDKRDASRKAE 9048
QY 179 ----- 178
Db 9049 WEEVNTAVEKRYGIWKLIIPNGQYEFVRVAVNKGISDECKSKVVIQDPYRLPGPGPKP 9108
QY 179 -----FNV 181
Db 9109 KVLARTKSMLYSWTPTPLDNGGSPITGYWLEKEEBSPYMSRVSRAPITKVGLKGVEFNV 9168
QY 182 PR----- 183
Db 9169 PRLLGKVYQFRAMAINAAGIGPSPSPDEVAGDPFPPPPSPCEVKDKTKSSISLGW 9228
QY 184 ----- 183
Db 9229 KPPAKDGGSPIKGYIVEMOEBGTDWKNVNEPKLITTCCEVVPNLKELKRYFRVKAVN 9288
QY 184 ----- 183
Db 9289 EAGESEPSDTTGBEIPATDIOEBEVEFIDIGAQDCLVCKAGSOIRIPAVIKGREPTPKSSWE 9348
QY 184 ----- 183
Db 9349 FDGAKKAMKDGVDHIDPEDAQLTAENSSVIIIEPCRSHTGKYSITAKNKAGQKTANCR 9408
QY 184 ----- 183
Db 9409 VKVMVPGPPKOLKVSDITRGCSRLSKWMPDDDDGGDRIKGYVIEKRTIDGKAWTKVNPDC 9468
QY 184 ----- 183
Db 9469 GSTTFVVPDLLSEQQYFFRVRAENRFGIGPPVETIORTTARDPIYPPDPPIKLKIGLITK 9528
QY 184 ----- 183
Db 9529 NTVHLSWKPPKNDGGSPVTHYIVECLAWDPTGTKEAWRQCNRKRDVEELQFTVEDLVEGG 9588
QY 184 ----- 183
Db 9589 EYEFVRKAVNAAGVSKSPATVGCDCQCPDMPSPSIDLKPFMEVEEGTNVNIKAKGVPF 9648
QY 184 ----- 183
Db 9649 PTLTWFKAPPKPDNKEPVLVYDTHVNKLVDVDTCLVI PQSRSDTGLYITITAVNNLGA 9708
QY 184 ----- 183
Db 9709 SKEMRLNVLGRPGPPVGPPIKFESVSADQMTLSWFPKDDGSGKITNVVIEKREANRKTW 9768
QY 184 ----- 183
Db 9769 HVSSEPKECTYTI PKLLEGHEVYFRIMAKNYKYGISEPLDSEPETARNLFSVPGAPDKPTV 9828
QY 184 ----- 183
Db 9829 SSVTRNSMTVNWEEPEYDGGSPVTGYWLENKDTTSKRWKRVNRDPIKAMTLGVSYKVTGL 9888
QY 184 ----- 183

Db	9889	IEGSDYQFRVYAINAAGVPSLPSDPATARDPIAPPGPPPKVDTWKSSADLEWSPL	9948	Db	10969	ECVVTGLQQGKTYFRVRAENIVGLGLPDTTPIEQEKLVPSPVELDVKLEGLVWKAG	11028
Qy	184	-----	183	Qy	208	-----	207
Db	9949	KDGSKVTVGYIVEYKEGKEWEKDKVTRGTLVVTGLKEGAFYKFRVSAVNIAGIGE	10008	Db	11029	TTVRFPAILIRGVPTAKWTTDGSSEIKTDEHYTVDNPFSSVLTIKNCLRRDTGEQIVT	11088
Qy	184	-----	183	Qy	208	-----	207
Db	10009	PGEVTDVIEWKDLVSPDLQLDASVRDRIVVHAGGVIRIIAIVSGKPPPTVWNNERTL	10068	Db	11089	SNAAGSKTVAVHLTVLDVPGPTGPINILDVTPHEMTISWQPPKDDGSPVINIYVEKOD	11148
Qy	184	-----	183	Qy	208	-----TALLRLYAQRNC-----	219
Db	10069	PQETIETTATSSMWIKNCORSHQGVYSLAKNEAGERKKTIIIVDVLDPGVPTPFLA	10128	Db	11149	TRKDTWGVSSGSKTKLPHLOKGEYVFRVRAENKIGVGPDLDTFTVAKHKEPSPS	11208
Qy	184	-----	183	Qy	220	-----	219
Db	10129	HNLTNESCKLTWSPEDDGGSPITNVVIEKRESREAWTPVYTVTRQNAATVQGLIQGA	10188	Db	11209	PPGKPVVTDITENAAATVSWTLPKSDGSSPITGYMERREVTGKWRVKNKTPIALDKFRVT	11268
Qy	184	-----	191	Qy	220	-----	219
Db	10189	YFFRIAAENSIGMGPFVETSEALVIREPIVPERPEDLEVKEVTKNTVLTWNPVKYDGG	10248	Db	11269	GLYEGNTYFRVFAENLAGLSKPSSDPIKACRPDKPPGPPINPKLKOKSRETADLVWT	11328
Qy	192	-----	191	Qy	220	-----	219
Db	10249	SEIINYVLESRLIGTEKFKVTDNDNLLSRKYTVKGLKEGTYEYRVSAVNIYQCKPSPFC	10308	Db	11329	KPLSDGSGPILGYVECOKPGTAGWNRINKDELIRQCAFVPGVGLIEGNEYFRPKAANIV	11388
Qy	192	-----	191	Qy	220	-----	219
Db	10309	TKPITCKDELAPTLHLDPRDKLTIRVGEAPALTGRYSGPKPKVSWFKDEADVLEDDRT	10368	Db	11389	GEGPRELAESVIAKDIHHPPEVELDVTCRDVITRVYGOTIRILARVKGRPEPDIWTKE	11448
Qy	192	-----	191	Qy	220	-----	219
Db	10369	HIKTPATLAEKIKAKRSDSGKYCVVVENSTGRKGFQCNVVDHPGPPGVPSFDEVT	10428	Db	11449	GKVLREKRVLDIQLPRVLEIQIKEAVRADHGKVIISAKNSSGHAQSAIVNVLDKPGPC	11508
Qy	192	-----	191	Qy	220	-----	219
Db	10429	KOYMWISWKPPLDDGGSKITNYIIIEKKEVGKDVWMPVTSASAKTCKVSKLLEKGDYIFR	10488	Db	11509	QNLKVTNVTKENCITSMENPLDNGSEITNFIVEYRKPNOKGSIVASDVTKRLIKANLL	11568
Qy	192	-----	191	Qy	220	-----	219
Db	10489	IHAENLYGISDPLVSDSMKADRFRVPDAPQPIVTEVTKDSALVTWKNPHDGGKPTNY	10548	Db	11569	ANNEYFRVCAENKVGVPPTIETKTPILAINPIDRPGEPENLHIADKGTFFVLKWRPD	11628
Qy	192	-----	191	Qy	220	-----	219
Db	10549	ILEKETMSKRWARTKDPHYPYKFRVPDLLEGQYEFVSAENEIGIDGSPSPKPVF	10608	Db	11629	YDGGPNLSYHVERBLKGSDDWVHKGSIKETHVMVDRCVENQIYEFVQTKNEGSESD	11688
Qy	192	-----	191	Qy	220	-----	219
Db	10609	AKDPIAKSPVNPPEAIDTTCNSVDLTWQPPRHDGSKILGYIYEQVKGDEWRRANHT	10668	Db	11689	WVKTEBWWVKEDLOKPLDLKLSGLTVKAGDTIRLEAGVRGKPPPEVAVTKDKDATDLT	11748
Qy	192	-----	191	Qy	220	-----	219
Db	10669	PBSCPETKYKVTGLRDGQTYKFRVLAVNAGESDPAHVPEPVLVKDRLEPPELILDANMA	10728	Db	11749	RSFRVKIDTRADSSKFSLTAKRSDGGKVVVTTATNTAGSFVAYATVNVLDKPGVPRNLKI	11808
Qy	192	-----	198	Qy	220	-----	219
Db	10729	REQHIKVGDTLRLSAILKGVPPPKVTWKEDRDAPTAKRIDVTPVGSKLEIRNAAHEDGG	10788	Db	11809	VDVSSDRCTVCWDPPEDDGGCEIQNYILEKCTKRMVWSTYSATVLTGTTVTRLIEGNE	11868
Qy	199	-----	198	Qy	220	-----	219
Db	10789	IYSLTVENPAGSKTVSVKVLVDKPGPRDLLEVSEIRKDSCYLTWKEPDLDDGGSVITNYV	10848	Db	11869	YIFRVAENKIGTPPTESKPIAKTKYDKPGRPDPPEVTKYVKEEMTVVWNPPEYDGGK	11928
Qy	199	-----	198	Qy	220	-----	219
Db	10849	VERRDASAQWSPLSATSKKSHFAKHLNEGNYLFRVAENQYGRGPPFVETPKPIKALD	10908	Db	11929	SITGYFLEKKEKHSRWPVNKSAIPERRMKVQNLDPDHEYQFRVAENEIGIGEPSLPS	11988
Qy	199	-----	198	Qy	220	-----	219
Db	10909	PLHPPGPKDLHHVDKTEVSLVWNNKPRDGGSPITGYLVEYQEGTQDWIKFTVTNL	10968	Db	11989	RPVVAKDPIBPPGPTNFRVVDTTKHSITLWGKPVYDGGAPITGYVEMRPKIADASPD	12048
Qy	199	-----	207	Qy	220	-----	219
		-----SQTIVFLGL-----		Db	12049	EGWKRCAAAQVLKREFTVTSLDENQEBYFRVCAQNVQVIGRPAELKEAIKPKLELPEPE	12108
		::: ::					

QY	220	-----	219	QY	234	-----	233
Db	12109	IDLDASMRKLVIVRAGCPIRLFAIVGRPAKVTWRKVGINVVRKQVLDVTWAPLVI	12168	Db	13189	DILIPPEGBHADLRKTLILRAGVTWRLYVPVKGRPPPKITWSPKNVNLDRIGLDIKST	13248
QY	220	-----	219	QY	234	-----	233
Db	12169	PNSTRDDSGKSYLTUNPAGEKAVFVNVRLDTPGVPVSDLKVSDDVTKTSCHVSWAPPEND	12228	Db	13249	DFDTPRCENVNKYDAGKYILTLENSCGKKEYTIVVKVLDTPGPPINVTVKESKDSAYV	13308
QY	220	-----TH-----	227	QY	234	-----	233
Db	12229	GGSQVTHYIVKREARKTWSVTVPVKKTSFHVTLVPGNEYYPRTAVNEYGPGVPTD	12288	Db	13309	TWEPPIIDGGSPIINVVQKRDARERKSWTVTTECSKTSFRVENLBEGKSYFFRVPFAENE	13368
QY	228	-----	227	QY	234	-----	233
Db	12289	VPKPVLASDPLSPDPKRLKLEATEMTKNGATLAWPLPLRDGGAKIDGYIISYREERQPAD	12348	Db	13369	YIGIDPGETRDVAKASQTPGPVVDLKVRSVSKSSCSIGMKPHSDGSGRIIGYVVDPLTE	13428
QY	228	-----	227	QY	234	-----	233
Db	12349	RWTEYSVVVDLSLVVTGLKEGKYKFRVAARNAGVSLPREAGVYEAKEQLLPKILMP	12408	Db	13429	ENKQVRVMSLSLOYSAKDLTSGKEYTFRVSAENENEGTPTSEITVVARDVVAPDLDLK	13488
QY	228	-----	227	QY	234	-----	233
Db	12409	EQITIKAGKLRTEAHVYGKPHPTCKWKKGEDEVVTSHLAVHKADSSILIIKDVTRKD	12468	Db	13489	GLPDLCLAKENSFRLKIPKGPAPSVSWKGEDPLATDTRVSVSSAVNTLLIVDC	13548
QY	228	-----	227	QY	234	-----	233
Db	12469	SGYSLTAENSSGTDQIKVVMADAPPPQPPDIDADACSLSMHIPLDGGSNIT	12528	Db	13549	QKSDAGKYITILKNVAGTKEGTISIKVWGKPGIPTGPIKFDEVTAEAMTLKWAPKDDGG	13608
QY	228	-----	227	QY	234	-----	233
Db	12529	NYIVEKDVSRGDWWTALASVTKTSRVGKLIPGQYIFRVAENRFGISEPLTSPQWVA	12588	Db	13609	SEITNYILEKRDVNNKWTCASAVQKTFRTVRLHEGMEYTFRVSANKNYGVGEGLKSE	13668
QY	228	-----	227	QY	234	-----	233
Db	12589	QPPPGVPSEPKNARVTKNKDCIFVADWRPDSGSGPIIGYLLIERKERNLLVWKANDTL	12648	Db	13669	PIVARHPPDVPDAPPPPNIVDVHRDSVSLTWDPKTGGSPITGYHLEBFKERNLLWKRA	13728
QY	228	-----	227	QY	234	-----PRV-----	236
Db	12649	VRSTYPCAGLVEGLEYSFRIYALNKAAGSPSPKTEYVTARMPDPPGKPEVIDTKST	12708	Db	13729	NKTPIMRDFKVTGLTEGLEYSFRVMAINLAGVGKPSLSEFPVALDPIPPGKPEVINI	13788
QY	228	-----	227	QY	237	-----PKYINGTKL-----	245
Db	12709	VSLIWARPKHGGSKIIGYFVEACKLPQDKWVRCNTAPHQIPQEBYTATGLEEKAQYQFR	12768	Db	13789	TRNSVTLLIWEPKYDGGHKLGTGYIVEKRDLPKSNWKAHVNVPECAFTVTDLVEGGKYE	13848
QY	228	-----	227	QY	246	-----	245
Db	12769	ATARTAVNISPPSESDPVTILAEVPPRIDLSVAMKSLTLTVKAGTNVCLDATVFGKPMF	12828	Db	13849	FRIRAKNTAGAISAPSESTETIICKDEYEAPTVILDPTIKDGITIKAGDTIVLNAISILG	13908
QY	228	-----AMGRNL-----	233	QY	246	-----	245
Db	12829	TVSWKKDGTLLKPAEGIKMAQORNLCTBLEFSVNRKDSGDYTTTAENSGSKSATIKLV	12888	Db	13909	KPLPKSSWSKAGKDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITITATNPFGTKVEHV	13968
QY	234	-----	233	QY	246	-----	245
Db	12889	LDKPGPPASVKINMYSDRAMLSWEPPLDGGSEITNYIVDKRETSRPNWAQVSATVPIT	12948	Db	13969	KVTVLDVPGPPGPVVISNVAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLMTVVSED	14028
QY	234	-----	233	QY	246	-----	246
Db	12949	SCSVEKLISGHEXQFRICAENKYGCDPVFTPEPAIKNPYDPPGRCDPPVISNITKDHMT	13008	Db	14029	IQSCRHVATKLIQGNEXIFRVSANVHYGKGPVQSEPVQWVDRFGPPGPKPEVSNVTK	14088
QY	234	-----	233	QY	247	NT-----MRKLK-----	253
Db	13009	VSWKPPADDGGSPITGYLLEKRETAQVNNTKVNRKPIIBRTLKATGLQGBTEYEFRTAI	13068	Db	14089	NTATVSWKRPVDDGGSEITGYHVERREKSLRWVRAIKTTPVSDLRCKVTGLQGBGYEPR	14148
QY	234	-----	233	QY	254	-----	253
Db	13069	NKAGPGKPSDASKAAYARDPQPPAPPAPPKVYDTRSSVSLSWGKPAYDGGSPIIGYLV	13128	Db	14149	VSAENRAGIGPPSEASDSVLKMDAAAYPPGPPSNPHVTDTTKKSASLANWKPHYDGGLEIT	14208
QY	234	-----	233	QY	254	-----	253
Db	13129	EVKRADSNWRCNLQNLQKTRFEVTGLMEDTQYQFRVAVNKNIGYSDPSDVPDKHYPK	13188	Db	14209	GYVVEHQVGDENAWIKDITGTALRITQFVVVPLQTKKYNFRISAINDAGVGEPAVDPV	14268
				QY	254	-----RKQAP-----	258

Db	14269	EIVEREMADPFELDAELRRLTVVRAGLSIRIFVPIKGRPAPEVTWTKONINLKNRANEN	14328
Qy	259	-----SYT-----	281
Db	14329	TSFTLLIPECNRYDTGKFVMTIENPAGKSGFVNRVLDTPGPVLNLRPTDITKDSVT	14388
Qy	259	-----	281
Db	14389	LHWDLPIDGSGRIITNYIIVEKREATRKSYSATTCKCHKTKYKVTGLSEGCYFFRVMAEN	14448
Qy	259	-----	281
Db	14449	EYIGICEPTTETBPVKASEAPSPDLSINIMDIITKTSVSLAWPKPKHGGSKITGYVIEAQR	14508
Qy	259	-----VKE-----	261
Db	14509	KGSDQWTHITTVKGLECVVRNLTEGBEYTFQVMAVNSAGRSAPRESRPVIVKEQTMLEP	14568
Qy	262	-----	261
Db	14569	DLRGIYQKLVIAKAGANIKEIPEVLGRPKFTVTKKGDQILKQTORVNEFTTATSTILNI	14628
Qy	262	-----	261
Db	14629	NECVRSDSGPYPLTARNIVGEVGDVITIQVHDIPGPPTGPIKPFDEVSSDFVTFSWDPDEN	14688
Qy	262	-----OPEKKA-----	267
Db	14689	DGGVPISNYVEMRQDSTTWVELATTVINTTYKATRLTGLLEYQPRVKAQNRYGVGPGI	14748
Qy	268	-----	267
Db	14749	TSAMIYANYPFKVPGPGTPOVTAVTKDSMTISWHEPLDGGSPILGYHVERKERNGILW	14808
Qy	268	-----	267
Db	14809	QTVSKALVPGNIFKSSGLTDGIAEYFRVIAENMAGSKSPKSEPMALDPIPPGKVPV	14868
Qy	268	-----	267
Db	14869	LNITRHTVTLKWAPEYTGFKITSYIIVEKRDLPNGRWLKANFNSILENEFTVSGLTEDA	14928
Qy	268	-----	267
Db	14929	AYEFRVIAKNAAGAI SPPSEPSDAITCRDDVEAPKIKVDVKFDTVLKAGEAPRLEADV	14988
Qy	268	-----	267
Db	14989	SGRPPTMWSKDGKELEGTALEIKIADFTNLVNRKSTRDSGAYTLTATNPGGFAXH	15048
Qy	268	-----	267
Db	15049	IFNVKVLDRPPEGPGPLAVTEVTSEKCVLSWFPPLDDGGAKIDHIYIVOKRETSRLATNV	15108
Qy	268	-----	267
Db	15109	ASEVQVTKLVTKLLKGNIEYIFRMAVNVKYGVEPLESEPVLAVNPYGPDPKPNPEVTT	15168
Qy	268	-----KKT-----	270
Db	15169	ITKDSMVVCGHPDSGGSEIINIYVERRDRAGORWIKCNKKTLDLRYKVSGLTEGHEY	15228
Qy	271	-----QSTTTPYF-----	278
Db	15229	BFRINAENAGISAPSTSPFYKACDVTFKPGPGNPRVLDTSRSSISIAWNKIYDGS	15288
Qy	279	-----	278
Db	15289	EITGTVVEIALPEEDEMQIVTPPAGLKATSVTITGLTENQBYKIRIYAMNSEGIGEPALV	15348
Qy	279	-----	278

Db	15349	PGTFKABDRMLPBEIELDADLRKVVITIRACCTLRLFPVIKGRPDPEVKWARDHGESLDKA	15408
Qy	279	-----SYT-----	281
Db	15409	SIESASVYTLIIIVGNVNRPDGKVIILTVENS SSGSAFVNRVLDTPGPQDLKVEVTK	15468
Qy	282	-----	281
Db	15469	TSVLTWDPPLLDGSGIKNYIIVEKRESTRKAYSTVATNCHTSKWSKVDQLQEGCSYFRV	15528
Qy	282	-----	281
Db	15529	LAENEYIGIGLPAETAESVKASERPLPGKITLMDVTNSVLSWEKPEHDGSGRILGYIV	15588
Qy	282	-----TSAALNVT-----	289
Db	15589	EMQTKGSDKWATCATVKVTEATITGLIQGBEYSPRVSQAONEKGISDPROLSVPVIAKDLV	15648
Qy	290	-----	289
Db	15649	IPPAFKLLFNTFTVLAGEDLKVDVFFIGRPTPAVTWHKDNVPLKQTRVNAESTENNSLL	15708
Qy	290	-----	289
Db	15709	TIKDACREDVGHYVVKLTNSAGEALETILNVILDKPGPPTGPKMDEVTA DSITLSWGP	15768
Qy	290	-----	289
Db	15769	KYDGGSSINNVIVEKRDSTTTTQIVSATVARTTIKACRLATGCEYQFRIAENRYGKST	15828
Qy	290	-----	289
Db	15829	YLNSEPTVAQYFKVPGPGPGPVVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERN	15888
Qy	290	-----	289
Db	15889	LWVKLNKTPIPQTKFTGLEEGVEYEFVRSAENIVGIGKPSKVSECYVARDPCDPPGRP	15948
Qy	290	-----	289
Db	15949	EAIIVTRNSVTLQWKPTYDGGSKITGYIVEKKELPEGRMMKASPTNIIDTHFEVTGLVE	16008
Qy	290	-----	289
Db	16009	DHRYEFRVIRARNAAGVFPSESESTGAITARDEVDPPIRISMDPKYKDTIIVWHAGESFKVDA	16068
Qy	290	-----	289
Db	16069	DIYGKPIPTIOWIKGQELSNTARLEIKSTDPATSLSVKDAVRVDSGNYILKAKNVAGER	16128
Qy	290	-----	289
Db	16129	SVTVNVKVLDRPPEGPGPVVISGVTAEBKCTLAWKPLQDGGSDIINYIVERRETSLVWT	16188
Qy	290	-----	289
Db	16189	VVDANVQTLSCVKYTKLLEGNEYTFRIMAVNKYGVEPLESEPVAKNPWPVDPAPKPEV	16248
Qy	290	-----	289
Db	16249	TTVTKDSMIWVERPASDGGSEILGYVLEKRDKEGIRWTRCHKRLIGELRLAVTGLIENH	16308
Qy	290	-----	289
Db	16309	DYEFVRSAENAGLSEPSFSAVQKACDPIYKPGPPNPKVIDITRSSVFLSWSKPIYDG	16368
Qy	290	-----	289
Db	16369	GCEIQGYIIVEKCDVNVGEWMTCTPPTGINKTNIEVEKLEKHEYNFRICAINKAGVGEHA	16428
Qy	290	-----	289
Db	16429	DVPGPIIVEEKLAPDIDLDELRLKINIRAGGSLRLFVPIKGRPTPEVKWKGVDGEIRD	16488

QY 290 ----- 289
Db 16489 AALIDVTSSFTSLVLDNVNRYDSKGYTLTLENSSGTSKSAFVTVRVLDTPSPVNLKVTEI 16548
QY 290 ----- 289
Db 16549 TKDSVITWEPPLLDGSGKIKNIVIEKREATRKSAAVVTNCHNSWKIDQLQEGCYYF 16608
QY 290 ----- 289
Db 16609 RVTAENEYIGLPAQTADPIKVAEVPQPKIITVDDVTNRNSVLSWTKPEHDGSGKIIQY 16668
QY 290 -----TNVT----- 293
Db 16669 IVEMQAKHSEKSECARVKSLOAVITNLTOGBEYLFVRVAVNEKGRSDRSLAVPIVAKD 16728
QY 294 ----- 293
Db 16729 LVIEPDVKPAFSYSVQVQDLMKEVPIISGRPKPTITWTXGDLPLKQTTTRINVTDSLDT 16788
QY 294 -----YSITTA----- 299
Db 16789 TLSIKETHKDDGQYGITVANVVGQKTASIEIITLDKPPKPVKFDVSAESITLSWN 16848
QY 300 ----- 299
Db 16849 PPLYTGCCQITNYIVOKRDTTITVMDVVSATVARTTLKVTCLKTGTGYOFRIPAENRYGQ 16908
QY 300 ----- 299
Db 16909 SFALSDPIVAQYPYKEPGPGTPFATAISKDSMVJQMHEPVNNGSPVIGYHLERKERN 16968
QY 300 ----- 299
Db 16969 SILWTKVNKTI IHDTOQAQNLLEGIEYEFVRVYAEINIVGVKASKNSECVYVARDPCDPPG 17028
QY 300 ----- 299
Db 17029 TPPEIMWGRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRMKNKFTNNVETQTVSGL 17088
QY 300 ----- 299
Db 17089 TEDQRYEFVRIAKNAAGAIKSPSDSTGPITAKDELVELPRISMDPKPRDTIWNAGTFRL 17148
QY 300 ----- 299
Db 17149 EADVHGKPLPTIEWLRGDKIEIESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAG 17208
QY 300 ----- 299
Db 17209 SKSFPVNVKVLDRPGPEGPVQVGTVSEKCSLTWSPLQDGGDISHYVVEKRETSRLA 17268
QY 300 ----- 299
Db 17269 WTVVASEVNTSLUKTLLEGNEYFRIMAVNKYGVGEPLSAFVLMKNPFLVLPKPKSL 17328
QY 300 -----ARRV----- 303
Db 17329 EVTNIADSWTCWNRPDSDGGSEIIGYIVEKDRSGIRWIKCNKRITDRLRVRTGLTE 17388
QY 304 -----STSTIAYR----- 311
Db 17389 DHEYEFVRSANAAGVGEPSAIVYKACDPVFKPGPPTNAHIVDTTKNSITLAWGKPIY 17448
QY 312 ----- 311
Db 17449 DGSSEILGYVVEICKADEEWQIVTQTLRLVTRFBISKLTEHQYKIRVCALNKVGLGE 17508
QY 312 ----- 311
Db 17509 ATSPGTVKPEDKLEAPELDLDELKRGIVVRAGGSARIHIPKGRPMPEITWSREGEF 17568

QY 312 ----- 311
Db 17569 TDKVOIEKGVNTQLSIDNCDNRDAGKYILKLENSSGSKSAFVTVRVLDTPGPQNLAVK 17628
QY 312 ----- 311
Db 17629 EVRKDSAFLWEPPIIDGGAKVKNYVIDKRESTRKAYANVSSKCKTSFKVENLTGAIY 17688
QY 312 ----- 311
Db 17689 YFRVMAENEFVGVVPVETVDVAKAAEPPPGKVTLTLDVTSQTSASLWKEKPEHDGGSRLV 17748
QY 312 ----- 311
Db 17749 GYVEMQPKGTEKWSIVAESKVCNAVVTGLSSGQEQYQFRVKAINEKGSQDPRVLGVFPVIA 17808
QY 312 ----- 311
Db 17809 KDLTIQPSLKLFPNTYSIQAGEDLKIEIPVIGRPRPNISWVKDGEPLKQTTRVNVSETAT 17868
QY 312 ----- 311
Db 17869 STVLHIKEGNKDDFGKYTVTATNSAGTATENLSVILEKPGPPVPRPDEVSADFWVIS 17928
QY 312 ----- 311
Db 17929 WBPPAYTGGCQISNVIWEKRDTTTTTHMVMSATVARTTIKITKLTGTGTQYQRIPAENRY 17988
QY 312 ----- 311
Db 17989 GKSAPLDSKAVIVQYPPFKPGPPGTPFTVTSISKQMLVQWHEBPVNDGGTKIIGYHLEQKE 18048
QY 312 ----- 311
Db 18049 KNSILWVNLKTIQDTKFTTGLDSELYEPKVSANIVIGIKPSKVSCEFVARPCDP 18108
QY 312 -----PDSSFMKS----- 319
Db 18109 PGRPEAIVTRNNVTLLWKPKPAYDGGKITGYIVEKKOLPDGRMKKASEFTNVLETFVS 18168
QY 320 ----- 319
Db 18169 GLVEDORYEFVRIARNAAGNFSEPSDSSGAIITARDEIDAPNASLDPKYKDVIVVHAGETF 18228
QY 320 ----- 319
Db 18229 VLEADIRGKPIPDVVWSKDGKLEBETAARMEIKSTIQKTTLLVVKDCIRTDGGQYILKLSN 18288
QY 320 ----- 319
Db 18289 VGGTKSIPITVKVLORPGSPGPKVTVGTABKCYLAMNPPLODGGANISHYIIEKRETS 18348
QY 320 -----IWA----- 322
Db 18349 RLSWTOVSTEQALNKYVKTLPLGNEYIFRVMANVKYIGEPLESGPVACPNKYPKPGPP 18408
QY 323 -----TQLR----- 326
Db 18409 STPEVSAITKDSNVVTVARPDVDDGGTEIESGYILEKEDKEGVRWTKCNKKTLDLRLRVGT 18468
QY 327 ----- 326
Db 18469 LTEGHSYEFRAAENAGVGEPSVPFYRACDALYPPGPPSNPKVTDTSRSSVSLAWSK 18528
QY 327 ----- 326
Db 18529 PIYDGGAPVKYVVEVKEAADEWTTCTPTGLQKQFTVTKLKENTEFNFRICAINSEG 18588
QY 327 ----- 326
Db 18589 VGEPATLPGSVVAQERIIEPPEIELDADLRKVVVLRASATLRLPVTIKGRPEPEVKEKAE 18648
QY 327 ----- 326

Db	18649	GILTRAOIEVTSFTMLVIDNTRFDSGRNLTLENNSGSKTAFVNVRLDSPAPVNL	18708
Qy	327	-----	326
Db	18709	TIREVKDSVTLSEWPEPLDGGAKITNVIEKRETRKAYATITNCKTTFRIENLOEG	18768
Qy	327	-----	326
Db	18769	CSYFVRLASNEYIGLPAETTEPVKVEPPLPPGRVTLVDVTRNTATIKWEKPESDGS	18828
Qy	327	-----	326
Db	18829	KITGVVEMQTKSEKSTCTQVKTLEATISGLTAGEEYVFRVAAVNEKGRSDPRLQVP	18888
Qy	327	-----	326
Db	18889	VIARDIEIKPSVELPHTFNVKARQOLKIDVPFKGRPOATVNRKDGQTLKETTRNVSS	18948
Qy	327	-----	326
Db	18949	SKVTLSLKEASKEDVGYELCVSNSAGSITVPIITIIILDRPGPPGIRIDEVSCDSIT	19008
Qy	327	-----	326
Db	19009	ISMNPPEYDGGCOISNYIVEKETTTHHIVSQAVARTSIKIVRLTTGSEYQFRVCAEN	19068
Qy	327	-----	326
Db	19069	RYGSSYSSSAVVAEYPPPGPPGTPKVHATKSTMLVTQVPVNDGSGRVIGYHLEY	19128
Qy	327	-----	326
Db	19129	KERSILWSKANKILLIADTVKVSGLDEGLMEYRVVAENIAGIKCSKCEVPAPDPC	19188
Qy	327	-----	326
Db	19189	DPGQPEVNTNRKSVLSKSPHYDGGAKITGYIVERRELPGDRLKCNVTNIQETYE	19248
Qy	327	-----	326
Db	19249	VTELTEDQRYEFRVFARNAADSVSEPSSTGPIIVKODVEPPRVMDVFRDVIIVKAGE	19308
Qy	327	-----	326
Db	19309	VLKINADIAGRPLPVISWAKDGIIEERARTEIISTDNHLLITVKDCIRRDGTQVYVTLK	19368
Qy	327	-----	326
Db	19369	NVAGTRSAVNCVKLDKPGPPAGPLEINGLTAEKCSLSWGRPOEDGGADIDYYHRKKRET	19428
Qy	327	-----	326
Db	19429	SHLAWTICBGLQMTCKVTKLLKGNEYIFRVTGVNKYGVGEPLSVAKALDFTVPSP	19488
Qy	327	-----	326
Db	19489	PTSLEITSVTKESNTLCSRPESDGGSEISGYIIEERKNSLRVVRNKKPVYDLRVKST	19548
Qy	327	-----	326
Db	19549	GLREGCEYRVAENAAGLSLPSLPIRAEDPVFLSPSPSKPIVDSGKTTITIAWV	19608
Qy	327	-----	326
Db	19609	KPLFDGAPITGYTVEYKSDDDTWKTSIQSLAGTEVTISGLTTGAEYVFRVSKVNKVG	19668
Qy	327	-----	326
Db	19669	SDPSDSSDPQIAKEREBEPLFDIDSEMRKTLIVKAGASFTMTVPFRGRVPFNVLSKPDPT	19728
Qy	327	DIAT-----	330
Db	19729	DLRTRAYVDTTDSRTSLATIENANRNDGKYTLTIQNVLASAASLTLVVKVLDTPGPPNIT	19788
Qy	331	-----	330
Db	19789	VQDVTKESAVLSWDVPENDGGAPVKNYHIEKREASKAWSVTNNCRLSYKVTNLQEGA	19848
Qy	331	-----	330
Db	19849	IYFRVSGENERGVIPIAETKEGVKITEKPSPEKLGVTISKDSVSUTWLKPEHDGSR	19908
Qy	331	-----	332
Db	19909	IVHYVVEALEKGQKNWVKCAVAKSTHVHVSGLRENSYFRRVFAENQAGLSDPRELLLPV	19968
Qy	333	-----	332
Db	19969	LIKEQLEPEIDMKNFPHSHVTVVRAGSNLKVDIPIISGKPLPKVTLSRDGVPLKATMRFT	20028
Qy	333	-----	332
Db	20029	EITAENLTINLKESVTADAGRYEITAANSSTGTTKAFINIVLDRPGPPGTVVISDITEE	20088
Qy	333	-----	332
Db	20089	SVTLKWEPPKVDGGSQVTVNYILLKRETSATAVTEVSATVARTMMKVMKLTGTGEEYOFRIK	20148
Qy	333	-----	335
Db	20149	AENRFGISDHIDSACTVTKLPYTPPGPSTPWTVNTVRESITVGNWHEPVSNGSAAVGVYH	20208
Qy	336	-----	335
Db	20209	LEMKDRNSILMQANKLVIRTHFKVTTISAGLIYEFVRVAENAAAGVGPSPSEPVLAI	20268
Qy	336	-----	335
Db	20269	DACEPRNRVITDISKNSVLSWQAPAPDGGSKITGYIVERRDLPDGRWTKASFTNVTET	20328
Qy	336	-----	339
Db	20329	QFTISGLTQNSQYEFVRVARNAGSISNPSEVGPITCIDSYGGPVIDLPLEYEVVKYR	20388
Qy	340	-----	339
Db	20389	AGTSVKLRAGISGKPAPTIWYKDDKELQTNALVCVENTTDLASILKADRLNSGGVEL	20448
Qy	340	-----	339
Db	20449	KLRNAMASATIRVOIILDKPGPGPIEFKTVTAEKITLLWRPPADDDGAKITHYIVEK	20508
Qy	340	-----	339
Db	20509	RETSRVVMSVSEHLEECIITTTKIKGNEYIFVRVAVNKYIGISPLESDSVAKNAFVT	20568
Qy	340	-----	339
Db	20569	PGPPGIPVTKITKNSMTVWVSRPIADGSDISGYFLEKRDKSLGWFKVXKETIRDRQ	20628
Qy	340	-----	347
Db	20629	KVTGLTNSDYQYRVCAVNAAGQPFSEFSEFYKAADPIDPPPPAKIRIADSTKSSITL	20688
Qy	348	-----	347
Db	20689	GWSKPVYDGSAGTVGVYVEIROGEBEETVSTKGEVTRTEYVVSNLKPGVNYFRVS	20748
Qy	348	-----	347
Db	20749	NCAGQGEPIENNEPVQAKDILEAPEIDLVALRTSVIAKAGEDVQVLIIPFKGRPPPTVTW	20808
Qy	348	-----	351
Db	20809	RKDEKLGSDARYSIENTDSSLLITIPQVTRNDTGKYLITLNGVGEPSSTVSVKVLDT	20868
Qy	348	-----	351

QY	352	-----	351	QY	364	-----	363
Db	20869	PAACQKLVHVSRTVTLMDPLDGGSPINYYVIEKRDATKRTWSVVSHKCSSTFK	20928	Db	21949	LDTPGPGPITPKDVTGRSATLMDAPLLDGGAR IHYYVVEKEASRRSQVISEKCTRQ	22008
QY	352	-----	351	QY	364	-----	363
Db	20929	LIDLSEKTPFFRVLAEBIGEPCETTEPVKAAEVPAPIRDLMSKDKTSVILSWTK	20988	Db	22009	IFKVNDLAAGVPYVFRVSAVNEYGVEPEPIVATQAPPPRLDVVDTSKSSAVLA	22068
QY	352	-----	351	QY	364	-----	363
Db	20989	POFDGGSVITEYVVERKGKEQTWSHAGISKTEIEVSQLEKQSVLEPRFAKNEKGLSD	21048	Db	22069	WLKPDHGGSRITGVLLMRQKSDLWEAGHTKQLTFTVERLVEKTEYEFVRVAKNDAG	22128
QY	352	-----	351	QY	364	-----	363
Db	21049	PVTIGPITVKELIITPEVDLSDIPGAQVTVRIGHNVHLELPYKPKPSIMLKQGLPLK	21108	Db	22129	YSEPREAFSIIKEPQIBPTADLTGITNQLITCKAGSPFTIDVPISGRPAPKPVTKLEE	22188
QY	352	-----	351	QY	364	-----	363
Db	21109	ESEFVRFSKTENKITLSIKNAKKEHGKVTVILDNAVCRIAVPITVITLGPSPKPGPIR	21168	Db	22189	MRLKETDRVSIITTKDRTTLTVKDSMRGSGRYFLTLENTAGVKTFSTVVVIGRPGVT	22248
QY	352	-----	351	QY	364	-----	363
Db	21169	FDEIKADSVILSHVDPEDNGGEBITCYIEKRETSQTNMWMVCSSVARTTFKVPNLVKDA	21228	Db	22249	GPIEVSSVSAESCVLSWGEPKDGGTEITNYIVEKRESGTTAWQLVNSSVKRTQIKVTHL	22308
QY	352	-----	351	QY	364	-----	370
Db	21229	EYQFVRAENRYGVSQLVSSIIIVAKHQFRIIPGPGPKVIYNTSDGMSLTWDAPVYDGG	21288	Db	22309	TKYMEYSFRVSSSENFRGVSKPLESAPIIAEHPFVPPSAPTRPEVYHVHSANAMSIRWEPY	22368
QY	352	-----	351	QY	371	-----	370
Db	21289	SEVTGFHVEKKERNILMKQVNTSPISGREYRATGLVGLDYQFVYAENSAGLSSPDP	21348	Db	22369	HDGGSKIIGYWVEKKERNITILMWKENKVPCLBCNYKVTGLVEGLEVOFRTYALNAAGVSK	22428
QY	352	-----	351	QY	371	-----	370
Db	21349	SKETLAVSPVDPGTPDYIDVTRITILKNPPLRDGGSKIYCYIEKQGNRWRCNP	21408	Db	22429	ASEASRPMAQNVPDAPGRPEVTDVTRSTVSLWSAPAYDGGSKVGVYIERKPVSEVD	22488
QY	352	-----	351	QY	371	-----	370
Db	21409	TDVSECQYTVGLSPGDRYEFRIIARNAVGTISPPSQSGIIMTRDENVPPIVEFGEYF	21468	Db	22489	GRWLKCNVTIVSDNFFVTVALSEGDTYEPRLAKNAAGVISKSESTGPVTCRDEYAPPK	22548
QY	352	-----	351	QY	371	-----	370
Db	21469	DGLIIKSGESLRKALVQGRPVPRVTWPKDGEIEKRMNMBITNVLGSTSLFVRDATRDH	21528	Db	22549	AELDARLHGLDVTIRAGSOLVLDAAVGGKPEPKIITWKGDKELDLCEKVSQYTKGRATA	22608
QY	352	-----	351	QY	371	-----	370
Db	21529	RGVYTVAKNAGSAAKAEIKVKQDTPGVGPPIRFTNITGEKNTLWMDAPLNDGCAPIT	21588	Db	22609	VIKFCDRSDSGKYTLTVKNASGTVKAVSVNVKVLDSFGPCGLTVSRVTOEKCTLAWSLPQ	22668
QY	352	-----	356	QY	371	-----	370
Db	21589	HYIEKRETSRLAWALLIEDKCEAQSATAIKLINGNEYQFVRSVAVNKFVGVRPLDSDPVVA	21648	Db	22669	EDGGAEITHYIVERRETSRLNWVIVEGECPTLSYVVVTRLIKNNEYIFRVRVAVNKYGGVP	22728
QY	357	-----	356	QY	371	-----	370
Db	21649	QIQYTVDPAGIPEPSNITGNSITLTWARPSDGGSEIOQYILEREKSTRWVKVISKR	21708	Db	22729	VESEPIVARNSTIPSPGPIPEEVGTGKEHIIQWTKPSDGGNEISNYLVDKREKESLR	22788
QY	357	-----	356	QY	371	-----	370
Db	21709	PISETRFKVTGLTEGNEYEFHVAENAGVGPASGISRLIKCREPNPPGPTVVKVTDIT	21768	Db	22789	WTRVNKDYYVVDTRLKVTLSMEGCDYQFRTAVNAAGNSEPSESNFISCREPSYTPGPP	22848
QY	357	-----	356	QY	371	-----	370
Db	21769	SKTTVSLEWSPKVPDGMELIIGYIEBMCKTDLDGWHKVAEACVKTRYTVTDLQAGEYK	21828	Db	22849	SAPRVDDTTKHSISLAWTKPMYDGGTDIVGYVLEMQEKDQWYVHTNATIRNTEFTVP	22908
QY	357	-----	363	QY	371	-----	370
Db	21829	FRVSAINGAGKGSCEVTGTIKAVDRLTAPELDIDANFKQTHVVRAGASIRLFYQGRP	21888	Db	22909	DLKMGQKYSFRVAANVKGMSYSSIAEIBPVERIEIPDLLEADDLKKTVTIRAGASLR	22968
QY	364	-----	363	QY	371	-----	370
Db	21889	TPTAVWSKPDNSLSLRADIIHTDTSFTLTVENCNRNDAGKYTLTVENNSGSKSITPTVKV	21948	Db	22969	LMVSVSGRPPPVITWTSKQIDILASRAIIDTTESYSLLIIVDKVNRVDAGKYTIEAENQSGK	23028
QY				QY	371	-----	370

Db 23029 KSATVLKVDYTPGPCPSVKVKEVSRDSTVIITWEIPTIDGGAPINNYIVEKREAAARAFK 23088
QY 371 -----TIYG----- 374
Db 23089 TWTTKSKTLYRISGLVEGTMHYFVRLPENIYIGCEPCTSDAVLSEVPLPAKLEVVD 23148
QY 375 ----- 374
Db 23149 VTKSTVTLAWKPLYDGGSLTGYLEACKAGTERMKVVTLKPTVLEHTVTSLEGEQY 23208
QY 375 -----TLDMSL----- 381
Db 23209 LFRIRAKNEKGVSEPRETVTAVTVQDLRLPTIDLSTMPQKTHVPAGRPVELVIPIAGR 23268
QY 382 ----- 381
Db 23269 PPPAASWFFAGSKLRESERVVETHTKVAKLTIRETTIRDTGEYTLKLNKVTGTTSETIK 23328
QY 382 ----- 381
Db 23329 VIILDKPPTGPIKIDEIDATISITISWEPPELDGGAPLSGYVVEQORDAHRPGWLVPSES 23388
QY 382 ----- 381
Db 23389 VTRSTFKFTRLTEGNEVFRVAATNRFGIGSYLQSEVIECRSSIRIPGPPETLIQIPDVR 23448
QY 382 ----- 381
Db 23449 DGMTLTWYPEDDGGQVTTYIIVERKEVRADRWVRNVKVPVTWRYRSTGLTEGLEYEHR 23508
QY 382 ----- 381
Db 23509 VTAINARGSKPRSPKPIVAMDPIAPPGKQPONPRVTDTRTSVSLAWSPEDEGGSKYT 23568
QY 382 ----- 381
Db 23569 GYLIEQKVQDHEWTKNTTPTKIREYTLTHLPQGAERPRVLACNAGGPGEPAEVPGTV 23628
QY 382 -----YNETMFV----- 389
Db 23629 KVTEMLEYDPYELDERYQEGIFVRQGGVIRLTPIKGKPPICKWYKEGQDISKRAMIAT 23688
QY 390 -----ENK----- 392
Db 23689 SETHTELVIKEADGDSGTVDLVLENKCGKAVVIKVRVIGSPNSPEGPLEYDDIQVRSV 23748
QY 393 ----- 392
Db 23749 RVSWRPADGGADILGILERRREVPKAAWYTIDSRVGTSLVVKGLKENVEYHFRVSAE 23808
QY 393 ----- 392
Db 23809 NQFGISKPLSEEPVTPKTLNPPPEPPNPPEVLDTVKSSVLSWSRPKDDGGSRVTGY 23868
QY 393 -----TASDS---NKT-----TPTSPS----- 406
Db 23869 IERKETSTDKVVRHNKTQIITWTYVTLVPDABEQFRIIAQNDVGLSETSPASEPVVCK 23928
QY 407 -----MGF----- 409
Db 23929 DPFDKPSQPGLEILSISKDSVTLQWEKPECDCGKEILGYWVEYRQSGDSAMKSNKERI 23988
QY 410 ----- 409
Db 23989 KDKOFTIGLLEATEYEFVRPAENETGLSRPRRTAMSIKTLTSGEAPGIRKEMKDVTK 24048
QY 410 ----- 409
Db 24049 LGEAAQLSCQIVGRPLDPIKWYRFGKELIQSRKYMSSDGRTHLTVMTBEQDEGVYTC 24108
QY 410 ----- 409

Db 24109 IATNEVGEVETSSKLLIQATPQFHPGPLKEKYYGAVGSTLRLHVMYIGRPVPAMTWFG 24168
QY 410 ----- 409
Db 24169 QKLLQNSENITIENTEHYTHLVMKNVQRKTHAGKVQLSNVFGTVDAILLDEIQDKPDK 24228
QY 410 ----- 409
Db 24229 PTGPVIEALLKNSAVISWKPADDGGSWITNYVVEKCEAKEGAEWQLVSSAISVTTCRI 24288
QY 410 -----ORTF--IDPL----- 417
Db 24289 VNLTENAGYYPRVAQNTFGISDFLEVSSVVIKSPPEKPGAPKPTITATKDCSCVAV 24348
QY 418 ----- 417
Db 24349 KPPASDGGAKIRNYLEKREKKQNKWISVTTEIRETVFSVKNLIEGLEYEFRVKCNLG 24408
QY 418 ----- 417
Db 24409 GESEWSEISEPITPKSDVPIQAPHFKEELRNLRNVYQSNATLVCKVTHGPKPIVKWYRQG 24468
QY 418 ----- 417
Db 24469 KEIIADGLKYRIOEPKGGHQLIITASVTDDBATVYQVRATNQGGSVSGTASLEVEVPAKI 24528
QY 418 ----- 417
Db 24529 HLPKTEGMAVHALRGEVWSIKIPFSGKPDVITWQGDLLDNNGHYQVIVTRSFSL 24588
QY 418 ----- 417
Db 24589 VFPNGVERKAGFYVVCANRFGIDQKTVELDADVDPDPGRGVKVSASRDSVNLWTWEP 24648
QY 418 ----- 417
Db 24649 ASDGSKITNYIVEKATTAERLWVGQARETRYVINLFGKTSYQFRVIAENKFLSKP 24708
QY 418 ----- 417
Db 24709 SBPSBPTITKEDKTRAMNVDREVDETVESMTKASHSTKELYEKYMAIEDLGRGEFGIV 24768
QY 418 ----- 417
Db 24769 HRCVETSKKTYMAKPVKVKGTDQVLVKEISILNIARHNILHLHESFESMEELVMIFE 24828
QY 418 ----- 417
Db 24829 FISGLDIFERINTSAFELNEREIVSVYHQVCEALQFLHSHNIGHFDIRPENIYYQTRRSS 24888
QY 418 -----W----- 418
Db 24889 TIKIIEFGQARQLKPDNFRLLFTAPEYYAVEPHQHDVVSTATDMMSLGTLVYVLLSGIN 24948
QY 419 -----DYLSLL----- 425
Db 24949 PFLAETNQOIENINMAEYTFDEEAFKEISIEAMDFVDELLVKEKSRMTASEALQHPWL 25008
QY 426 ----- 425
Db 25009 KQKIERVSTKVIRTLKHRRYVHTLIKDLNMVVSNAARISCGGAIRSQKGVSAKVKVASI 25068
QY 426 -----FLDEIRNF----- 433
Db 25069 EIGPVSQIMHVBEGGHVKYCVKIENYDQSTQVWTWYFGVRLNSEKEYEITYEDGVAI 25128
QY 434 ----- 433
Db 25129 LYVKDITKLDGTYRCKVNDYGEDSSYAEFLVKGVREYVYYCRRTMKKIKERTDTMRL 25188
QY 434 ----- 433
Db 25189 LERPPFTLPLYNKTAAYGVNVRFGVTITVHPHPHTWYKSGQIKPGDNDKKYTFESDK 25248

Qy	434	-----	433
Db	25249	GLYQLTINSVTTDDDAEYTVVARNKYGEDSCKAKLTVTLHPPTDSTILRPMFKRLANAE	25308
Qy	434	-----	433
Db	25309	QBGQSVCFEIRVSGIPPTLKWKDGGQLSLGPNIBIIHEGLLOYALHIDTLPEDTGY	25368
Qy	434	-----	433
Db	25369	YRVTAINTAGSTSCQAHLQVERLRYKKQBFKSEHERHVQKIDKTLRMAEILSGTESV	25428
Qy	434	-----	433
Db	25429	PLTQVAKREALREAAVLKPAVSTKTVKGBFRLEIBEKKEERKLMPYDVPBRKYKQTTI	25488
Qy	434	-----	433
Db	25489	EEDORIKQFVPMDSMKYKIRDOYEMPGKLDVVQKPKRILRLSRWEQFVYVMPLPRI TD	25548
Qy	434	-----	433
Db	25549	QYRPKWRI PKLSQDDLEIVPARRRTPSPDYFYRPRRRSLGDISDELLLPIDDYLAM	25608
Qy	434	-----SLR----- 	436
Db	25609	KRTEERLRLBEELELFGASPSPSRKPPHFELSLRYSYPOAHKVEETRKNFRYSTYHI	25668
Qy	437	-----	436
Db	25669	PTYAEASTYAEULRERHAQAAYRQPKORIMAREDEBELLRPVTTTQHLSEYKSELDFM	25728
Qy	437	-----SPTYVNL----- 	443
Db	25729	SKEEKSKRRQREYTEITEEVEEISKHAQRESSSASALLRRRSLSTYIELMRP	25788
Qy	444	-----TP----- 	445
Db	25789	VSELIRSPQPAEYEDDTERSPPTERTPRSPSPVSEKSLRNFERSARFDIFSYES	25848
Qy	446	-----	445
Db	25849	MKAALTKQTSERKEYEVLSQLPTLDDHAPRILRMRSHRVPQGNTRFILNVQSKPTAEV	25908
Qy	446	-----	445
Db	25909	KWYHNGVELQESSKIHYTNTSGVLTLEILDCHDDSGTYRAVCTNYKGEASDYATLDVTG	25968
Qy	446	-----	445
Db	25969	GDYTTYASORRDEEVPSPVPELTRTEAYAVPSFKTSEMEASSSVREKVSQMTETRESL	26028
Qy	446	-----	445
Db	26029	SSYEHSASAEKSALEEKSLBEKSTTRKIKTTLAARILTKPRSMVYVEGESARFSCD TD	26088
Qy	446	-----	445
Db	26089	GEPVPTVTLRKQVLSTARSQVTTTKYKSTFEISSVQASDEGNYSVVVENSEKQBAE	26148
Qy	446	-----PEHRAV----- 	452
Db	26149	FTLTIOKARVTEKAVTSPPRVKSPBRVKSPBAVKSPRKRKVSPPSPHKA VSPPTETKPTP	26208
Qy	453	-----	452
Db	26209	REKVQHLPVSA PKITQPLKAEASKEIAKLTCVVESSVLRAKEYTWYKDGKKLXENGHPQ	26268
Qy	453	-----NLSGN----- 	458
Db	26269	PHYSADGTVELKINNLNLTESDQGEYVCEISGEGGTSKTNLQFMGAQFASIHFKVSKISETK	26328

QY	459	-----SLW 461
Db	26329	KSDQKTTSTVTRKTEPKAPEPISSKPFVIVTGLQDITVSSDSSVAKFAVKATGEPRTAIW 26388 :
QY	462	----- 461
Db	26389	TKDGAITOGGKYKLSEDKGGFLEIHKHTDTSDSGLYCTTVKNSAGSVSSCKLTIKAIK 26448
QY	462	----- 461
Db	26449	DTEAQKVSTQKTSBITPQKAVVQBEISQALRSEBIKNSAKSQEKLALKEEASKVLIS 26508
QY	462	----- 461
Db	26509	EEVKSAATSLEKSIVHEBITKTSQASEEVRTHAEIKAFSTQMSINEGORLVLKANIAGA 26568
QY	462	---W----- 462
Db	26569	TDVKWLNGVELTNSSEYRYGVSGDQTITIKQASHRDEGILTCISKTKEGIVKCYDLT 26628
QY	463	-----WLQ 465 :
Db	26629	LSKELSDAPAFISQPRSQNINEQNVLFTCEISGEPSPEIEWEK 26672
RESULT 5		
ID	AAU05396	
XX	AAU05396 standard; protein; 26926 AA.	
AC	AAU05396;	
DT	24-OCT-2001 (first entry)	
XX	Human titin (connectin) protein sequence.	
XX	Human; titin; connectin; pickwick mutation; cardiac specific exon; N2B;	
KW	titin-related disease; zebrafish; heart failure; heart disease.	
XX	Homo sapiens.	
XX	WO200151666-A1.	
XX	19-JUL-2001.	
PD	12-JAN-2001; 2001WO-US001212.	
XX	12-JAN-2000; 2000US-0175787P.	
PR	(GEO) GEN HOSPITAL CORP.	
XX	Fishman MC;	
PI	WPI; 2001-451869/48.	
PI	N-PSDB; AAS05390.	
DR	Determining if a subject has or is at risk of developing a titin-related	
PT	disease or condition, particularly heart failures, comprises detecting	
PT	the presence of a mutation in the titin gene.	
XX	Disclosure; Page 57-111; 114pp; English.	
PS	The present sequence representing human titin (also known as connectin)	
XX	is described in an invention relating to a novel method for determining	
CC	whether a subject has or is at risk of developing a titin-related disease	
CC	or condition. The method comprises analysing a nucleic acid of a sample	
CC	from the subject and detecting the presence of a mutation (e.g. the	
CC	pickwick mutation in the cardiac specific exon N2B) in the titin gene,	
CC	which indicates that the subject has or is at risk of developing a titin-	
CC	related disease. The zebrafish which has a phenotype similar to mammalian	
CC	heart failure is used as a model. The method is useful for detecting an	
CC	increased likelihood of heart disease, such as heart failure, in a	
CC	patient, so that appropriate intervention can be instituted before any	
CC	symptoms occur. The method may also be used to facilitate determination	

CC	of etiology of an existing heart condition, such as heart failure, to									
CC	identify compounds that can be used to treat or prevent heart conditions,									
CC	in prenatal genetic screening, e.g. to identify parents who may be									
CC	carriers of a recessive titin mutation. Compounds identified using the									
CC	methods may be used to treat patients that have or are at risk of									
CC	developing heart disease, e.g. heart failure									
XX										
SQ	Sequence 26926 AA;									
	Query Match	27.7%; Score 678; DB 4; Length 26926;								
	Best Local Similarity	1.0%; Pred. No. 1.5e-06;								
	Matches 267; Conservative	93; Mismatches 105; Indels 25559; Gaps 85;								
QY	1	MGRK-----	4							
		: :								
Db	649	VGKKAERAVTVAAVDQARVREPPEFHLEESYAQQTILEYGYKERISAAKVAEPPQRA	708							
QY	5	-----	4							
Db	709	SEPHVVPKAVPRVIAQPSETHIKTTDQGMHISSQIKKTTDLTTLRLVHVDKRPETASP	768							
QY	5	-----	4							
Db	769	HFTVSKISVPKTEHGYEASIAIATLOKELSATSSAQKITKSVKAPTVPKSETRVRAE	828							
QY	5	-----	4							
Db	829	PTPLPOFPADPTDYKSEAGVEVKVGVSITGTTVREERFVLHGREAKVTETARVPA	888							
QY	5	-----	4							
Db	889	PVEIPVTPPTLVSLGNVTVIEGESVTLECHISGYPSTVWYREDYQIESSIDFQITFQ	948							
QY	5	EMVRD-----	10							
		: :								
Db	949	SGIARLMIREAFADSGRFTCSAVNEAGTVSTCYLAVQVSEFEKETTAATEKFTTEEK	1008							
QY	11	-----	10							
Db	1009	RFVESRDVMTDSLTEEQAGPEPAAPYFITKPVVQKLVEGGSVFGVCGVGNPKPHVY	1068							
QY	11	-----	10							
Db	1069	WKSGVPLTGYRYKVSYNKQTGECKLIVISMTFADDAGEYTIIVRNKHGETSASALLEE	1128							
QY	11	-----	10							
Db	1129	ADVELLMKSQOEMLYQTQVTAFOEPEVGETAPGFVYSEYEKEQEALIRKMAKDTV	1188							
QY	11	-----	10							
Db	1189	VVRTYVEDQEFHISSFEERLIKEIYRIIKTTLEELLEEDGEKMAVDISESEAVESGFD	1248							
QY	11	-----VPK-----	13							
		:								
Db	1249	LRIKNVRILEGMVTFHCKMSGYPLPKIAWYKDGRIKHGERYQMDFLQDGRASLRIPVV	1308							
QY	14	-----	13							
Db	1309	LPEDEGIYTFASNIKGNAICSGKLYVEPAAPLGAPTYIPTLEPVSRIKSLSPRSVSRSP	1368							
QY	14	-----MFVLISISF-----	22							
		: :								
Db	1369	IRMSPARMSPARMSPARMSPGRRLEETDESQLERLYKPVFVLKPVSKFLEQOTA	1428							
QY	23	-----	22							
Db	1429	RFDLKVVGRPMPTFWFHGQQIIVNDYTHKVIKEDGTQSLIIVPATPDSGSEWTVVAQN	1488							
QY	23	-----	22							
Db	1489	RAGRSSIVLTVEAVEHGVKPMFVEKLVNVIKESRLEMKVRATGNPNPDIVLWLNKSD	1548							

QY	23	-----	22							
DB	1549	IIVPHKYPKIRIEGTGKGEALKIDSTVSQDSAWYTATATINKAGRDTTTRCKVNVVEFAEP	1608							
QY	23	LLV	-----	25						
DB	1609	EPERKLIIPRGTYRAKEIAAPELPLHLRYQOEOWEGDLYDKEKQKPPFKKLTSLRL	1668							
QY	26	-----	SF	27						
DB	1669	KRFGPAHPECLRTIPISDPTMVVWMLHDGKPLEAANRLRMINEFGVCSLDYGVAYSRDGSI	1728							
QY	28	INCKVMSK	-----	35						
DB	1729	ITCRATNKYGTHTSATLIVKDEKSLVEESQLPEGRKGLQRIEELERMAHEGALTGVTTD	1788							
QY	36	ALNRPWR	-----	43						
DB	1789	QKEKQKPDIVLYPEPVRVLEGETARFCRVGTGYPQPKVNVYINGQLIRKSRFRVRYDGI	1848							
QY	44	GLVLSKI	-----	50						
DB	1849	HYLDIVDCKSYDTGEVKVTAENPEGVIEHKVLEIQOQREDFRSVLRRAPEPRPEPHVEP	1908							
QY	51	-----	50							
DB	1909	GKLOPEVQKVRPVDTTTETKEVVKLKRAERITHEKVPESSEBELSKFKRTEEGYEAIT	1968							
QY	51	-----	50							
DB	1969	AVELKSRKKDESYPELLRKTDELLHWTKELTTEEKKALAEKGKTIITFKPKDIELSPS	2028							
QY	51	-----	50							
DB	2029	MEAPKIFERIQSTVGQSDAHFRVVRVVGKPDPECEWYKNGVKIERSDRIYWPEDNVC	2088							
QY	51	-----	50							
DB	2089	ELVIRDVTAEDSASIMVKAINIAGETSSHAFLVQAKQILITFTQELQDVVAKEKDTMATF	2148							
QY	51	-----	50							
DB	2149	ECETSEPPFVKVWYKDGMEVHEGDKYRMSDRKHFLSLITLITDSDAEDYSCVLVEDENV	2208							
QY	51	-----	GKYL	55						
DB	2209	KTTAKLIVEGAVVEFKELQIEVPESYSGELECIVSPENIEGKWHYHNDVELKSNKXTI	2268							
QY	56	-----	DOLK	59						
DB	2269	TSRRGRQNLTVKDVTKEDQGEYSFVIDGKTKTKLKMFRPIAILQGLSDQKVCEDIVQ	2328							
QY	60	LEILRQLET	-----	70						
DB	2329	LEVKSLESVEGWKMGQGEVQPSDRVHIVIDQSHMLLIEDMTKEDAGNYSFTIPALGL	2388							
QY	71	-----	70							
DB	2389	STSGRVSVYSDVITPLKDVNVIEGTAVLECKVSPDVTSVKWLNDQIKPDRVQAI	2448							
QY	71	-----	70							
DB	2449	VKGTQORLVINRTHASDEGPYKLVGRVETNCNLSVEKIKIRGLRDLTCTETQNVFV	2508							
QY	71	STKNYS	-----	77						
DB	2509	ELSHSGIDVLWNFKDKEIKPSSKYKIEAHGIYKLVLMNMKDDGKTYFYAGENMTSGK	2568							
QY	78	-----	77							
DB	2569	LTVAGGAISKPLTDQTVAESQEAFFECVNPDPKGEWLKGLPLTNNIRSESDGHR	2628							
QY	78	-----	KQPVKNLTM	86						

Db 2629 RLIIAATKLDIGEYTYKVATSKTSAKLVKVAVKIKTKLNLFTVTQDAVFTVELTHEN 2688
QY 87 -----: : : : : :----- 86
Db 2689 VKGVQWIKNGVWLESNEKVAISVKGTTIYSLRIKNCALVDESUYGFRGLRGASARLHVET 2748
QY 87 -----: : : : : :----- 86
Db 2749 VKIIKPKDVTALENATVAFESVSHDTVPVKWFHKSVEIKPSDKHRLVSEKRVHKLMLQ 2808
QY 87 -----NTERPO----- 92
Db 2809 NISPSDAGEYTAVGQLECKAKL FVETLHITKTKMKNIEVPETKTASFECEVSHFNVPMSW 2868
QY 93 -----: : : : : :----- 92
Db 2869 LKNGVEIEMSEKPIVVQGLHQLIIMNTSTEDSAEYTFVCGNDQVSATLTVTPIMISM 2928
QY 93 -----: : : : : :----- 92
Db 2929 LKDJNAEEDTITFEVTNVYEGISYKWLKNGVEIKSTDKQMRTKKLTHSLNIRNVHFGD 2988
QY 93 -----: : : : : :----- 92
Db 2989 ADYTFVAKATSTATLYVEARHIEPRKHIDIKVLEKRAMFECEVSEPDITVQWKKDD 3048
QY 93 -----YYILAG----- 98
Db 3049 QELQITDRIKIQEKYVHRLIIPSTRMSDAGKTVVAGGNVSTAKLFVSGRDRVIRSIKK 3108
QY 99 -----: : : : : :----- 98
Db 3109 EVQVIEKORAVFEVNEDDVAHWYKDGIEINFQVQERHKYVWERRIHRMFISETRQSD 3168
QY 99 -----: : : : : :----- 98
Db 3169 AGEYTFVAGNRSSVTLYNAPPEPOVLOELQPVTVQSGKPARFCAMISGRPOPKISWYK 3228
QY 99 -----: : : : : :----- 98
Db 3229 BEQLLSTGPKRFLHDGQEVYLLLLIEAFPEDAAYTCEAKNDYGVATTASLSVEVPEV 3288
QY 99 -----PION----- 102
Db 3289 SPDOEMPVYPPAIITPLQDVTVSEGQPARFCQVSGTDLKVSWSKDKKIKPSRPFMTQ 3348
QY 103 -----: : : : : :----- 102
Db 3349 FEDTYQLEIAEAYPEDEGTYTFVANNVQVGSSTANLSLEAPESILHERIEQEIEMEMKE 3408
QY 103 -----: : : : : :----- 102
Db 3409 FSSSFLSABEGLHSAELQSKINETLELLSSEVPVYTFKDESEKGTGPIFIKEVSNADI 3468
QY 103 -----: : : : : :----- 102
Db 3469 SMGDVATLSVTVIGIPKPIQWFFNGVLLTPSADYKVFVFDGDDHSLIILFTKLEDBEY 3528
QY 103 -----: : : : : :----- 102
Db 3529 CMASNDYKTI CSAYLKINSKGECHKDTETESAVAKSLEKLGCGPCPPHFLKELKPIRCAQ 3588
QY 103 -----YSL-----TYLWF----- 110
Db 3589 GUPAIFEYTVGEPATVTFWFKENKQLCTSVVYTIHNPNGSGTFIVNDPQREDSGLYIC 3648
QY 111 -----: : : : : :-----DFYSTQLRKPA----- 121
Db 3649 KAENMLGESTCAEALLVLLDMDTDTPCAKSTPEAPDFQTPKLGPAVEALDSEQEI 3708
QY 122 -----: : : : : :----- 121

Db 3709 ATPVKDITILKAALITEENQOLSVEHIAKANELSSQLPLGAQELQSLILEODKLTPESTRFP 3768
QY 122 -----: : : : : :----- 121
Db 3769 LCINGSIHFQPLKEPSNQLQIIVSQKTFKSKEGILMPPEPETQAVLSDETEKIFPSAMSI 3828
QY 122 -----: : : : : :----- 121
Db 3829 EQINSLTVEPLKTLAABPEGNYQSSIBPPMHSYLTSAEVLSLKERTVSDTNRQORVT 3888
QY 122 -----: : : : : :----- 121
Db 3889 LQKQEAQSALILSLSLAEGHVESLOSQPDVMIQVNYEPLVPSEHSCTEGKILIESANPL 3948
QY 122 -----: : : : : :----- 121
Db 3949 ENAGQDSAVRIEKGSLRFPALAEKQVLLKEHSDNVVMPDQIIESKREPVAIKKVOE 4008
QY 122 -----: : : : : :----- 121
Db 4009 VQGRDLLSKESLLSGIPEQRLNLKIQICRALQAVASEQPLFSEWLRNIEKVEYAVN 4068
QY 122 -----: : : : : :----- 121
Db 4069 ITQEPRHIMCYLVTSAKSVTEVTIIIEDVDPQMANLKMELRDALCALIYBEIDILTAE 4128
QY 122 -----: : : : : :----- 121
Db 4129 GPRIQQAKTSLQEEMDSPSGQKVEPIPEPEVESKYLISTEBSYFNQSRVKYLDATP 4188
QY 122 -----: : : : : :----- 121
Db 4189 VTKGVASAVVDEKQESLKPSEKSESSESGTBEVATVKIQEABGGLIKEDGPMIHTP 4248
QY 122 -----: : : : : :----- 121
Db 4249 LVDTVSEGDIVHLTTSITNAKEVNWYFENKLVPSDEKFKCLQDQNTYTLVIDKVTEDH 4308
QY 122 --KYVYSQYNTHTAKTI----- 135
Db 4309 QGEYVCEALNDSGKTATSAKLTWVKRAAPVIRKRIEPLVALGHLAKFTCEIQSAPNVR 4368
QY 136 -----: : : : : :----- 135
Db 4369 QWFKAGREIYESDKSIRSKEYISSLEILRTQVDCGYTCKASNEYSVGSCTATLTIV 4428
QY 136 -----: : : : : :----- 135
Db 4429 PGGEKKVRKLLPERKPEKPEEVVLSVLRKRPEEBEPKVEPKKLEKVKPAVPEPPPKP 4488
QY 136 -----TFRPPP----- 141
Db 4489 VBEVEVPTVKERKIPEPTKYPEIKPAIPALPAPKPKPEAEVKTIKPPVPEPEPTPIA 4548
QY 142 -----: : : : : :----- 141
Db 4549 APVTVPVVGKKAABAKAPKEAAKPGPIKGVPKTSPSIEABRRKLRPGSGGKPPDEAP 4608
QY 142 -----: : : : : :----- 141
Db 4609 FTYQLKAVPLKFVKKEIKDIIILTESEBFVGSAAIFECVLSPSTAITTWMKOGSNIRESFKUR 4668
QY 142 -----: : : : : :----- 141
Db 4669 FIADGKDRKLHIIDVQLSDAGEYTCVLRGNKEKTSTAKLVVEELPVRPVKTLLEBEVTV 4728
QY 142 -----C-----GRV----- 145
Db 4729 KGQPLYLSCELNKBRDVVVRKDGKIWEKPGRIVPQVIGLFRALTINDADTDAGTYTVT 4788
QY 146 -----: : : : : :----- 145
Db 4789 VENANNLECCSVCVKVVEVIRDLWLVKPIRQHVKPGTAFACDIAKDTNIRKPKGYDEI 4848

QY 146 ----- 145
Db 4849 PAEPNDKTEILRDGNHLYLKIKNAMPEDIAEYAVIEGKRYPAKLTGGEREVELLKPIED 4908
QY 146 -----PSMTC----- 150
Db 4909 VTIYEKESAFDAIESEADIPGOWKLKGELLRPSPTCEIKABGGKRFLLTHVKVLDQAGE 4968
QY 151 ----- 150
Db 4969 VLYQALNAITTAILTVKEIELDFAVPLKDVTPERRQARFECVLTREANVINSGPDIK 5028
QY 151 ----- 150
Db 5029 SSDKFDIIADGKHILVINDSQDDDEGVTAIEVEGKTSARLFTVGTIRLKFMSPLEDQTV 5088
QY 151 ----- 150
Db 5089 KEGETATFVCELSHERKHVVWPKNDAKLHTSRVTLISSEBQTHKLEMKEVTLDDISQIKA 5148
QY 151 ----- 150
Db 5149 QNKELSSTAQLKYLEADPYTVKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVSP 5208
QY 151 ----- 150
Db 5209 KYSIKADGLRLILKIKADLKDKGEVVCDCGTDKTKANVTVEARLIEVEKPLYGVEFVG 5268
QY 151 ----- 150
Db 5269 ETAHFEIELSEPDVHGOWKLKGQPLTASPDCEIIEDGKKHILHNCQLGWTGEVSQAA 5328
QY 151 ----- 150
Db 5329 NAKSAANLKVKEPLIFITPLSDVKVPEKDEAKFECEVSEBPKTFRWLKGTOBITGDDRF 5388
QY 151 ----- 150
Db 5389 ELIKDGTKHSNVIKSAFDEAKMPEAKEDKHTSGKLIIBGIRLKFTPLKDVTAKEKES 5448
QY 151 ----- 150
Db 5449 AVFTVELSHDNIRVKWPKNDQRLHTRSVSMQDEGKTHSITPKDLSIDDTSQIRVEAMGM 5508
QY 151 ----- 150
Db 5509 SSEAKLTVLESDPYFTGKLODYTGVEKDEVILQCEISKADAPVKWPKDGEIKEPSKNAVI 5568
QY 151 ----- 150
Db 5569 KTDGKKEMLILKKALKSDIGQYTCDCGTDKTSGLKDIEDREIKLVRPLHSVEVMEETAR 5628
QY 151 ----- 150
Db 5629 FETEISDDIHANWKLKGEALLOTPDCEIKEEGKIHSLVHLNCRLDQGTGVPQAANVKS 5688
QY 151 ----- 150
Db 5689 SAHLRVKPRVIGLLRPLKDVTVTAGETATPDCELSYEDIPIVEWYLGKKLEPSPDKVPR 5748
QY 151 ----- 150
Db 5749 EGKVHTILTRDVKLEDAQEVQLTAKDFKTHANLPVKEPPVEFTKPLEDQTVBEGATAVLE 5808
QY 151 ----- 150
Db 5809 CEVSRENAKVKWFKNGTEILSKKYEIVADGRVRLKVIHDCPTEDIKTYTCDAKDKFTSC 5868
QY 151 ----- 150
Db 5869 NLNVVPHVEFLRPLTLQVREKEMARFECELSRENAKVKWFKDGEIKKKGYDIIISKG 5928

QY 151 ----- 150
Db 5929 AVRILVINKCLLDDAEYSCVETARTSGMLTVLEEEAVFTKNLANIEVSETDTIKLVCE 5988
QY 151 ----- 150
Db 5989 VSKPAEVIWYKGBDEIETGRYIELTEGRKRIILVIONAHLEDAGNYNCRPLSPSRTDGKV 6048
QY 151 ----- 150
Db 6049 KVHELAAEFISKPNLEILEGEKAEFVCSISKESFPVQWKRDDKTLES GDKYDVIADGKK 6108
QY 151 ----- 150
Db 6109 RVLVVKOATLQDMGTYYVMVGAAARAAHLTVIEKLRIVVPLKDFRVKESQEVVFNCEVNT 6168
QY 151 ----- 150
Db 6169 EGAKAKWPRNEAIPDSSKYIILQKDLVYTIIRDALHDDQANYVSLTNHGENVKSA 6228
QY 151 ----- 150
Db 6229 NLIIVEEDLRIVEPLKDIETMEKKSVTFCWKVNLNVLTKWTONGEEVFPDNRSYRVDK 6288
QY 151 ----- 150
Db 6289 YKHLMTIKDCGFPDRGEYIVTAGODKSVAEILLIIEAPTEFVEHLEBDQVTFEFDDAVFCQ 6348
QY 151 ----- 150
Db 6349 LSREKANVWYRNGREIKEGKYKFEKDGSHRLLIKDCLDDCEYACGVEDRKSARL 6408
QY 151 -----LSEM----- 154
Db 6409 FVEEIPVEIIRPPQDILEAPGADVVFVLAELNKDKVEQVQLRNNMNVVQGDQHMSEGKI 6468
QY 155 ----- 154
Db 6469 HRLQICDKPRDQGEYRFIAKDKEARAKLELAAPKIKTADODLVVDVKPLTMVVPYDA 6528
QY 155 ----- 154
Db 6529 YPKAAEWFKENEPLSTKTIDTTABQTSFRILEAKKGDKGRYKIVLQNKHGKAEGLNLK 6588
QY 155 ----- 154
Db 6589 VIDVPGVRNLEVTETFDGEVSLAWEEPLTDGSKIIGYVVERDDIKRKTWYLATDRAES 6648
QY 155 -----LNV----- 157
Db 6649 CEFTVTGLQGGVEYLFVRSARNRVGTGEPVETDNPVEARSKYDVPGPLNTIITDVNRF 6708
QY 158 ----- 157
Db 6709 GVSLTWEPPEYDGAETINYVIELRDKTSIRWDTAMTVRAEDLSATVTDVBEQYESFRV 6768
QY 158 ----- 157
Db 6769 RAQNRIGVGKSAATPFVKVADPIERPSPVNLTSDDQTSQSVQLKWEPLKDGSGPILG 6828
QY 158 ----- 157
Db 6829 YIIERCEGKNWIRCNMKLVPELTYKVTGLEKGNKYLVRSAENKAGVSDPSEILGPLT 6888
QY 158 ----- 157
Db 6889 ADDAFVEPTMDLSAFKDGLEIVPNPITILVPSTGYPRPTATWCFGDKVLETDGDRVKMT 6948
QY 158 ----- 157
Db 6949 LSAYABLVISPSERSDKIYTLKLENVKITISGEIDVNVVIARPSAPKELKFGDITKDSVH 7008
QY 158 ----- 157

Db	7009	LWEPDDGGSLTGYYVEKREVSRTWTKVMDFTVDTLEFVPLVQGGKEYLKVCAARN	7068	Db	8089	VEVHNPTAEAMTITWKPPPLYDGGSKIMGYIIIEKIAKGBERWKRCNEHLVPILTYTAKGLE	8148
Qy	158	-----	157	Qy	170	-----	169
Db	7069	KCGGEPAYVDFPNMSTPATVPDPENVKWRDRTANSIFLTWDPKNDGSGRIKGYIVE	7128	Db	8149	EGKEYQFRVAENAAGISEPSRATPPTKAVDPIDAPKVILRTSLEVKRGDEIADASISG	8208
Qy	158	-----	157	Qy	170	-----	169
Db	7129	RCPRGSKWVACGEPVAETRMVEVTGLEEGKWAYRVKTLNRQGASKSPRTEBIAQVDTQ	7188	Db	8209	SPYPTITWIKDENVIIVPEEIKKRAAPLVRRRKGEVQEBBPFVLPLTQRLSIDNSKKGESQ	8268
Qy	158	-----	157	Qy	170	-----	169
Db	7189	EAPFIFLDVKLAGLTVKAGTKIELPATVTGKPEPKITWTKADMILKQDKRIITIENVPKK	7248	Db	8269	LVRDSLRLPDHGLYMIKVENDHGIAPACTVSVLDTPGPPINPFVFDIRKTSVLCKWEPP	8328
Qy	158	-----	164	Qy	170	-----	172
Db	7249	STVTIVDSKESDGTGYIIIEAVNVGRATAVVENVLDKCPAPAAFDITDVNBSCLLTWN	7308	Db	8329	LDDGGSEIINYLTLEKKDKTKPDSEMIWVTSLRHKVSVTKLIEGKEYLFRVAENRFGP	8388
Qy	165	-----	164	Qy	173	-----	178
Db	7309	PPRDDGSKIYVVERRATDSBWMHKLSTVKDTNFKATKLI PNKEYIFRVAENNYGA	7368	Db	8389	GPPECVSKPLVAKDPFGPPDAPDKPIVEDVTSNMLVKWNEPKDNGSPILGYMLEKEVNS	8448
Qy	165	-----	164	Qy	179	-----	178
Db	7369	GEVQASPIAKYQFDPGPPTRLEPSDITKDAVTLTWCEPDDGGSPITGYWVERLDPD	7428	Db	8449	THWSRVNKSLLNALKANVDLLEGLTYVFRVCAENAAAGPKFSPSPDPKTAHDPISPPGP	8508
Qy	165	-----	164	Qy	179	-----	178
Db	7429	TDKWRCNKMPVKDYYRVKGLTNKKYRFRVLAENLAGPKSPKSTEPILIKDIPDPW	7488	Db	8509	PIPRVTDTSSTTIELEWEPPAFNGGGEIVGYFVDQLVGTNKNWSRCTEKMIKVQYTVKE	8568
Qy	165	-----	164	Qy	179	-----	178
Db	7489	PPCKPTVKDVGKTSVRLNMTKPEHDGAKIESVVIEMLKTGTDEWVRVAEGVPTTQHLLP	7548	Db	8569	IREGADYKLRVSAVNAAGEGPPGETQPVTVABPQEPAPVELDVSKGGIIMAGKTLRIP	8628
Qy	165	-----	164	Qy	179	-----	178
Db	7549	GLMEGOEYFRVAVNAKAGESESPSPVLCREKLYPPSPRWLEVINITKNADLKW	7608	Db	8629	AVTGRPVPTKVWTKBEGELDKORVIDNVGKSELIIKDALRKHGRVYVITATNSCGSK	8688
Qy	165	-----	164	Qy	179	-----	178
Db	7609	VPEKGGSPITNIVKRDVRRKGWQTVDTTVKDTCTVPLTEGSLYVFRVAENAIQ	7668	Db	8689	FAARVEFVDPGVLDLKPVVTRNKMCLLNWSDEPDGSGEITGPIIERKOKAMHTWRQ	8748
Qy	165	-----	164	Qy	179	-----	178
Db	7669	SDYTEIEDSVLAKDTFTTGGPYALAVDVTKRVDLKWEPKNDGGRPIQRVIEKKER	7728	Db	8749	PIETERSKCDITCLLEGGQYKFRVIAKNKFGCGPPVEIGPILAVDPLGPTSPERLTYTE	8808
Qy	165	-----	169	Qy	179	-----	178
Db	7729	LGTRWVAKGTAGDCNFRVTDVIEGTEVQFQVRAENEAGVGHPSPTILSIEDPTSP	7788	Db	8809	RORSTITLDWKEPRNGSGSPIQGYIIIEKRRHDKDPFERNVKNLCPPTSPLENLDEHMY	8868
Qy	170	-----	169	Qy	179	-----	178
Db	7789	SPPLDLHTDAGRKHIAIAWKPEKNGGSPIIGYHVEMCPVTEKMRVNSRPIKDLKPK	7848	Db	8869	EFRVKAVNEIGSEPSPLNVVIOQDEVPPTIKLRLSVRGDTIKVKAGEPVHVPADVTGL	8928
Qy	170	-----	169	Qy	179	-----	178
Db	7849	VEBGVVPDKEYVLRVAVNAIGVSESEISENVVAKDPCKPTI DLETHDII VIEGEKLS	7908	Db	8929	PMPKIWSKNETVIEKPTDALQITKEEVSRSBAKTELSIPKAVREDKGTGTVTASNRLGS	8988
Qy	170	-----	169	Qy	179	-----	178
Db	7909	IPVFRVAVPVTVSMHKGCKEVRASDLTMKNNDHISAHLEVPKSVRADAGIYITILENKL	7968	Db	8989	VFRNVHVEYDRPSPPRNLAVTDIKAESCYLTWDAPLNDGSGSEITHYVIDKRDASRKKAE	9048
Qy	170	-----	169	Qy	179	-----	178
Db	7969	GSATASINVKVI GLPGCKDIKASDITKSSCKLTWEPPEFDGTPILHYVLERREAGRRT	8028	Db	9049	WEVNTTAVEKRYGIWKLI PNQOYEFVRVAVNKYGISDECKSKVKVIOQPYRLPGPPGKP	9108
Qy	170	-----	169	Qy	179	-----	181
Db	8029	YIPVMSGNKL SMTVKDLIPNGEYFFFRVAVNVKVGGEVIELKNPVIAQDPKQPPDPVD	8088	Db	9109	KVLARTKGSMLVSWTTPLDNGGSPITGYWLEKREBGS PYWSRVSRAPITKVGLKGVEFNV	9168
Qy	170	-----	169	Qy	182	PR-----	183
Db				Db	9169	PRLEGVKYQFRAMAINAAGIGPPSEPDPVAGDPFPPPPSPCEVKDKTKSSISLGW	9228

QY 184 ----- 183
 Db 9229 KPPAKDGGSPKIGYIVEMOEGTDMKRVNEPDKLITTCBVCVPNLKLKRYFRVAVN 9289
 QY 184 ----- 183
 Db 9289 EAGESEPSDTGEIPATDIOEBPEVIDIGAODCLVCKAGSQRIPAVIKGRPTPKSSWE 9348
 QY 184 ----- 183
 Db 9349 FDGKAKKAMKGVHDIPEDAQLEAENSSVIIPECKRSHTKGYSITAKNKAGQKTANCR 9408
 QY 184 ----- 183
 Db 9409 VKYMDVPGPPKOLKVSDITRGSCLRSWKMPDDGGDRIKGYVIEKRTIDGKAWTKVNPDC 9468
 QY 184 ----- 183
 Db 9469 GSTTFVVPDLLSEQOYFFRVRAENRFGIGPPVETIQRTTARDPIYPPDPPIKIKIGLITK 9528
 QY 184 ----- 183
 Db 9529 NTVHLSWKPPKNDGSPVTHYIIVECLAWDPTGTKEARQCNRDVEELOFTVEDLVEGG 9588
 QY 184 ----- 183
 Db 9589 EYEFVRKANNAAGVSKPSATVGPDCQCPDMPSPSIDLKEFMEVEEGTNVNIIVAKIKGVPPF 9648
 QY 184 ----- 183
 Db 9649 PTLTWFKAPKPDNKEPVLVDTHVNKLVDVDDCTLVIPSQRRSDTGLYITITAVNNLGT 9708
 QY 184 ----- 183
 Db 9709 SKEMRLNVLRGPPVGPPIKESVSADQMTLSWPPKDDGSKIITVYIEKREANRKTWV 9768
 QY 184 ----- 183
 Db 9769 HVSSEPKECTYTIIPKLLGHEHYFRIMAQNKYIGIBPLDSEPETARNLFSVPGAPDKPTV 9828
 QY 184 ----- 183
 Db 9829 SSVTRNSMTVWEEPEYDGGSPVTGYWLEMDTTSKRWKRVNRDPIKAMTLGVSYKTGL 9888
 QY 184 ----- 183
 Db 9889 IEGSDYQFRVYAINAAGVPASLPSPDPATARDPIAPGPPPPKVTWTKSSADLEWSPPL 9948
 QY 184 ----- 183
 Db 9949 KDGSKVGTGYIVEYKEGKEWEKGDKEVGTGLVVTGLKEGAFYKFRVSAVNIAGIGE 10008
 QY 184 ----- 183
 Db 10009 PGEVTDVIEMKDLVSPDLQDASVRDRIVVHAGGVIRIIAYVSGRPPPTVTWNMERTL 10068
 QY 184 ----- 183
 Db 10069 PQEATIEHTAISSSWVIKNCORSHQGVYSLAKNEAGERKTIIVDVLDPVPGVGPFLA 10128
 QY 184 ----- 183
 Db 10129 HNLTNESCKLTWFSPEDDGGSPITNYVIEKRESRRRAWTPVTYVTRQONATVOGLIQGKA 10188
 QY 184 ----- 191
 Db 10189 YFFRIAAENSIGMGPFVETSEALVIREPITVPERPEDLEVKEVTNKTVTTLTWNPKYDGG 10248
 QY 192 ----- 191
 Db 10249 SEIINYVLESRLIGTEKHFHKTNDNLLSRKYTVKLGEGDTEYVRVSAVNIIVGQKPSFC 10308

QY 192 ----- 191
 Db 10309 TKPITCKDELAPPTLHLDFRDKLTIRVGEAFALTGRYSGKPKVKVSWFKDEADVLEDDRT 10368
 QY 192 ----- 191
 Db 10369 HIKTTPATLALAKIKAKRSDSGKYCVVVENSTGSRKGFQVNVVDHGPVPPVGPVSFDEVT 10428
 QY 192 ----- 191
 Db 10429 KDMVVISWKPLDDGGSKIITNYIIIEKKEGVDMVMPVTSASAKTTCKVSKLLEGKDIYER 10488
 QY 192 ----- 191
 Db 10489 IHAENLYGISDPLVSDLSMAKADRFVPDAPQPIVTEVTKDSALVTWNKPHDGGKPIITNY 10548
 QY 192 ----- 191
 Db 10549 ILEKRETMKSKWARVTKDPIHPYTKFRVPDILLLEGQCYEFRVSAENEIGIDGPPSPKPVF 10608
 QY 192 ----- 191
 Db 10609 AKDPIAKPSPVNPBEADITTCNSVDLTWQPPRHDGSKILGYIVEYQKVGDDEWRANHT 10668
 QY 192 ----- 191
 Db 10669 PESCPETKYKVTGLRDGQTYKFRVLAVNAAGESDPAHVPEPVLVKDLRLEPPELILDANMA 10728
 QY 192 ----- 198
 Db 10729 REQHIKVGDTLRLSNIKGVPPFPKVTWKEDDAPTAKRIDVTPVGSKLEIRNAAHEDGG 10788
 QY 199 ----- 198
 Db 10789 IYSLTVENPAGSKTVSVKVLVDKPGPRDLEVSEIRKDSCLTWKPELDDGGSVITNYV 10848
 QY 199 ----- 198
 Db 10849 VERRDVASAQMSPLSATSKKSHFAHLNEGQYLFVAAENQYGRGPFVETPKFIKALD 10908
 QY 199 ----- 198
 Db 10909 PLHPPGPKDLHHVDVDKTEVSLVWKNKPRDGGSPITGYLVEYQEBGTQDWIKFKVTNVL 10968
 QY 199 ----- 207
 Db 10969 ECVWTLQOGKTYFRFRVKAENIVGLGLPDTTPIECQEKLVPPSVELDVKLEGLVVRAG 11028
 QY 208 ----- 207
 Db 11029 TTVRFPALIRGVVPPTAKWTTDGSBIKTDEHYTVETDNFSSVLTIKNCLRRDTGEVQITV 11088
 QY 208 ----- 207
 Db 11089 SNAAGSKTVAVHLTVLDVPGPTGPIINILDVTPHEMTISWQPPKDDGGSPVINYIVEKOD 11148
 QY 208 ----- 219
 Db 11149 TRKDTMGVVSSGSSKTKLPHLQKGEYVFRVRAENKIGVGPPLDSTPTVAKHKFSPPS 11208
 QY 220 ----- 219
 Db 11209 PFGKPVVTDITENAAVSWTLPKSDGSGSPITGYMYMERREVTGKWRVKNKPTIADLKFRVT 11268
 QY 220 ----- 219
 Db 11269 GLYEGNTYFRVFAENLAGLSKSPSSDPIKACRPIKPPGPPINPKLKDOKSRETADLVWT 11328
 QY 220 ----- 219
 Db 11329 KPLSDGSGSPILGVVVEQKPGTAQWNRINKDELIROCAPRVPLIEGNEYFRIKAANIV 11388
 QY 220 ----- 219

Db 11389 GECEPRELAESVIAKDILHPPPEVELDVTCRDVITVRVGQITIRILARVKGRRPBDITWTKE 11448
QY 220 ----- 219
Db 11449 GKVLVREKRVLDLIQDLPRVELQIKEAVRADHGKYIISAKNSSGHAQSAIVNVLDPRGPC 11508
QY 220 ----- 219
Db 11509 QNLKVTNVTNKENCTISWENPLDNGSGSEITNFIVEYRKPQKQGSIVASDVTKRLIKANLL 11568
QY 220 ----- 219
Db 11569 ANNEYFRVCAENKVGVPETIETKTPIILAINPIDRPGEPENLHDKGTFFVYLKWRPD 11628
QY 220 ----- 219
Db 11629 YDGGSPNLSYHVERRLKGDDMERVHKGSIKETHYVMDRCVENQIYEFPRVQTKNEGSED 11688
QY 220 ----- 219
Db 11689 WVKTEBWWVKEDIQKPVLDIKLSGVLTVRKAGDITRLEAGVRGKPPPEVAVMTKDKOATDLT 11748
QY 220 ----- 219
Db 11749 RSPRVKIDTRADSSKPSLTKAKRSDGGKYVVTATNTAGSFVAVATVNVLDKPGFVRNLKI 11808
QY 220 ----- 219
Db 11809 VDSSDRCTVCWDPPDDGCEIQNYILEKCEKRMVWSTYSATVLTPGTTVTRLEGNE 11868
QY 220 ----- 219
Db 11869 YIFRVAENKIGTPTESKPIAKTKYDKGPRDPPEVTVKSKBMTVMVNPPEYDGGK 11928
QY 220 ----- 219
Db 11929 SITGYFLEKKEKSTRVWPVNKSAIPERRMKVONLLPDHEYQPRVKAENEIGIBPSLPS 11988
QY 220 ----- 219
Db 11989 RPVWAKDPIEPGPPNFRVVDVTKHSITLWGKPVYDGGAPFIYGVVEMRPKIADASPD 12048
QY 220 ----- 219
Db 12049 EGWKRCAAAQLVRKEFTVTSLDENQYEFVRVCAQNVGIGRPAELKEAIKPKILEPPE 12108
QY 220 ----- 219
Db 12109 IDLDASMRKLVIVRAGCPIRLFAIVRGRPAKPVTRKVGIDNVVRKGQVDLVDTWAPLVI 12168
QY 220 ----- 219
Db 12169 PNSTRDSSKYSLTLVNPAGEKAVFVNVRLDTPGVPVSLKVSDDVTKTSCHVSWAPPEND 12228
QY 220 ----- 227
Db 12229 GGSQVTHYIVEKREADRKTWSTVTPSVKKTSPHVTNLVPGNEYFRVTAVNEVPGVPTD 12288
QY 228 ----- 227
Db 12289 VPKPVLASDPLSEBPPRKLKLEATEMTKNGATLAWPLPLRDGGAKIDGYIISYREEBPAD 12348
QY 228 ----- 227
Db 12349 RWTEYSVVXDSLVLVTGLKEGKYKFRVAARNVAVGSLPREAGVYEAKEQLLPPKILMP 12408
QY 228 ----- 227
Db 12409 EQITIKAGKRLIEAHVYKPHPTCKWKKGEDBVTSSHLAVHKADSSSILIKDVTRKD 12468
QY 228 ----- 227

Db 12469 SGYISLTAENSGTDTQIKIVVMDAPGPPQPPFDISDIDADACSLSWHIPLEGGSNIT 12528
QY 228 ----- 227
Db 12529 NVIVEKDVSRGDMWTALASVTKTSRUGKLIPIQGEYIIFRVAENRFGISEPLTSPKWA 12588
QY 228 ----- 227
Db 12589 QPFPFVGPSPKNARVTKVNKDCIFVWDRPDSGGSPPIGYLIERKERNLLWVKANDTL 12648
QY 228 ----- 227
Db 12649 VRSTYPCAGLVEGLEYSFRIYALNKAAGSPSPKTEYVTARMPVDPCKPBEVIDVTKST 12708
QY 228 ----- 227
Db 12709 VSLIWARPKHDGSKIIGYFVEACKLPGDKWVRCNTAPHQIQEEYATATGLEBAQYQFR 12768
QY 228 ----- 227
Db 12769 AIARTAVNISPPSEPDVPTILAENVPRIDLSVAMKSLITVKAGTNVCLDATVFGKMP 12828
QY 228 ----- 233
Db 12829 TVSWKDKGTLLKPAEGIKWAMORNLCLELFSVNRKDSGDIYITAEENSGSKSATIKLV 12888
QY 234 ----- 233
Db 12889 LDKPGPASVINKMYSRAMLSWBPLEDGGSEITNYIVDKRETSRPNWAQVSATVPIT 12948
QY 234 ----- 233
Db 12949 SCSEKLIIEGHEYQFRIKAENKYGVDPVFTBPAIAKNPYDPGRCDPPVISNITKDHT 13008
QY 234 ----- 233
Db 13009 VSWKPPADGGSPITGYLLEKRETOAVNWKVNRKPIERTLKATGLQGBTEYEFRTAI 13068
QY 234 ----- 233
Db 13069 NKAGPKPSDASKAAVARDPQYPPAPPAPKVVYDTRSSVSLWGKPAYDGGSPIIIGYL 13128
QY 234 ----- 233
Db 13129 EVKRADSNWVRNCLPONLQKTRPEVTGLMEDTQYQFRVAVNKIGYSDPSVDPRHYPK 13188
QY 234 ----- 233
Db 13189 DILIPPEGHDADLRKTLILRAGVTMRLYVPVKGRPPPKITWSKPNVNLDRIGLDIKST 13248
QY 234 ----- 233
Db 13249 DFDTLRCENVNKYDAGKYILLTLENSCGKKEYTIVVKVLDTPGPPINVTVKEISKDSAV 13308
QY 234 ----- 233
Db 13309 TWEPIIDGGSPIINVVQKRDAERKSWSTVTTECKTSFRVPLEEGSKSYFRVFAENE 13368
QY 234 ----- 233
Db 13369 YGIGDPGETRDAVKASQTPGPVVDLKVRSVSKSGSIGWKPHSGGSRIIIGYVVDFLTE 13428
QY 234 ----- 233
Db 13429 ENKQVRVMSLSLOYSAKDLTEGKEYTFRVSAENENGCTPSEITTVARDDVVVAPDLDL 13488
QY 234 ----- 233
Db 13489 GLPDLCLAKENSFRLKIPKGPAPSVWKKGBDPLATDTRVSESSAVNTTLIVYDC 13548
QY 234 ----- 233
Db 13549 QKSDAGKVTITLKNVAGTKEGTISIKVVGKPGIPTGPIKFDEVTAEAMTLKWA PKDDGG 13608

QY	234	-----	233	QY	262	-----	267
Db	13609	SBITNYLEKRDSDVNNKWTCASAVQKTTFRVTRLHEGMEYTFRVAENKYGVBGLKSE	13668	Db	14689	DGGVPISNVYVEMRQDSTTWELATTVIRTYYKATRLTTLGLEQFRVKAQNRYGVGPGI	14748
QY	234	-----	233	QY	268	-----	267
Db	13669	PIVARHPDVPDAPPPPNIVDRHDSVSLTWDPKTKGSPITGYHLEPKERNLLWKRA	13728	Db	14749	TSAWIVANYPKVPGPPTQVTAVKDSMTISWHBPLSDGGSPILGYHVERKERNGILW	14808
QY	234	-----	236	QY	268	-----	267
Db	13729	NKTPIRMDFKVTGLTEGLEFRVMAINLAGVGKPSLPBPVALDPIDPFGKEVINI	13788	Db	14809	QTVSKALVPGNIPKSSGLTDGIAEFVRIAENMAGKSPKSEPMALALDIPDPGPVP	14868
QY	237	-----	245	QY	268	-----	267
Db	13789	TRNSVTLIWTEPKYDGGHKLGTGYIVEKRDLPKSMKANHVNPCECAFTVTDLVEGGKYE	13848	Db	14869	LNITRHTVTLKWAKPEVTGGFKITSYIVEKRDLPNGRWLKNFNSILNEFTVSGLTEDA	14928
QY	246	-----	245	QY	268	-----	267
Db	13849	FRIRAKNTAGIASPSESTETIICKDEYEAPTIVLDPITIKDGLTIKAGDTIVLNAISILG	13908	Db	14929	AYEFRVIAKNAAGAISSPSEPSDAITCRDDVEAPKIKVDVKFKDTVILKAGEAFRLADV	14988
QY	246	-----	245	QY	268	-----	267
Db	13909	KPLPKSSWKAGKDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITATNPFGTKVHV	13968	Db	14989	SGRPPPTMEWSDGKLEGTAKLEIKIADFSTNLVNKDSRTRDSGAYTLTATNPGGFAKH	15048
QY	246	-----	245	QY	268	-----	267
Db	13969	KVTLDVPGPGPVEISNVSAEKATLTWTPPLEDGGSPIKSYILEKRETSRLLWTVVSSED	14028	Db	15049	IFNVKVLDRPGPBGPLAVTEVTSEKCVLSWFPPLDDGGAKIDHYIVQKRETSRLATNV	15108
QY	246	-----	K 246	QY	268	-----	267
Db	14029	IQSCRHVATKLIQNEVI FRVSAVNHVGKEPVQSEPVKMWDRFGPPGPEKPEVSNVTK	14088	Db	15109	ASEVOVTKLKVTLLKMGNEYIFRVMAVNVKYGVEPSEPVIAVNPYPPDPKPNVTT	15168
QY	247	NT-----	253	QY	268	-----	270
Db	14089	NTATVSKRPVDDGGSEITGYHVERBKSLRWRAIKTPVSLRCKVTGLQEGSYEPR	14148	Db	15169	ITKDSMVVWCHGPDSDGSEIINVIYERRDKAGORWIKCNKKTLLTDLRYKVSGLTEGHEY	15228
QY	254	-----	253	QY	271	-----	278
Db	14149	VSAENRAGICPPSEASDSVLMKDAAYPPGPPSNPHVDTTKKSASLAWGKPHYDGGLEIT	14208	Db	15229	EFRIMAENAGISAPSPSPFYKACDVTFKPGPPGNPRVLTDSRSSISIAWNKPIYDGG	15288
QY	254	-----	253	QY	279	-----	278
Db	14209	GYVVEHQVGEAWIKOTTGTALRITQFVVVDLQTKENYFRISAINDAGVGEPAVDPV	14268	Db	15289	EITGYMVEIALPEDEWQIVTPPAGLKATSYITIGLTENQEVKIRIYAMNSEGLGEPAV	15348
QY	254	-----	258	QY	279	-----	278
Db	14269	EIVEREMAPDFELDAELRRTLIVRAGLSIRIFVPIKGRPAPEVTTWKDNLNKRANEN	14328	Db	15349	PGTPKAEDRMPLPPIELDADLRKVVTIRACCTLRLFVPIKGRPDPEVKWARDHGESLDA	15408
QY	259	-----	258	QY	279	-----	281
Db	14329	TESFTLLIIPECNRYDTGKFVMTIENPAGKSGFVNVRLDTPGPVNLNRPDTDKDSVT	14388	Db	15409	SIESASSYTLIIVGNVNRFDGKYILTVENSNGSKSAFVNVRLDTPGPQDCLKVEYTK	15468
QY	259	-----	258	QY	282	-----	281
Db	14389	LHWDPLIDGSRITNYIVEKREATRKSYSTATTCKHCTYKVTGLSEGCEYFRVMAEN	14448	Db	15469	TSVTLTWDPPLDDGSKIKNYIVEKRESTRKAYSTVATNCHTKSWKVQLOBGCSYFVR	15528
QY	259	-----	258	QY	282	-----	281
Db	14449	EYIGEPTEPTEPVKASEAPSPDLSNIMDITKSTVSLAWPKPKHDGSKITGYVIEAQR	14508	Db	15529	LAENBYGIGLPAETAESVKASERPLPGKITILMDVTRNSVLSWBKPEHDGSRILGYIV	15588
QY	259	-----	261	QY	282	-----	289
Db	14509	KGSDQWTHITTVKGLCVVRNLTEGEEYTFQVMAVNSAGRSAPRESRPVIVKEQTMLEP	14568	Db	15589	EMQTKGDKWATCATVKKVTEATITGLIOGEEYSFRVSAQNEKGISDPRQLSVPVIAKDLV	15648
QY	262	-----	261	QY	290	-----	289
Db	14569	DLRGYQKLVIAKAGDNIKVEIPVLRPKPTVTWTKGQDILKQTORVNFETTATSTILNI	14628	Db	15649	IPPAFKLLFNTFTVLAGEDLKVDPFPIGRPTPAVTWHKONVPLKQTVRNAESTENNSLL	15708
QY	262	-----	261	QY	290	-----	289
Db	14629	NECVRSDSGPPLTARNIVGEVGDVITIQVHDIFGPPTGPIKPFDEVSSDFVTTSWDPPEN	14688	Db	15709	TIKDACREDVGHVVKLVNSAGEALETNLVLDKPGPPTGVKMDVETADSIITLSWGP	15768
				QY	290	-----	289

Db	15769	KYDGGSSINNYIVEKRDSTTTWQIVSATVARTTIKACRLKTCGEYQFRIAAENRYGKST	15828	Db	16849	PPLYTGCGCOITWYIVQKRDTTTTWDVVSATVARTTLKVTCLKTGTGYQFRIIPAENRYGQ	16908
QY	290	-----	289	QY	300	-----	299
Db	15829	YLNSEPTVAQYPFKVPGPCTPVVTLSSRDSMEVQWNEPISDGGSRVIGYHLERKERN	15888	Db	16909	SFALESDDPIVAQYPYKEPGPGTPPATAISKDSMWIQWHEPVNNGSPVIGYHLERKERN	16968
QY	290	-----	289	QY	300	-----	299
Db	15889	LWVKNLKTPIQTKFTTGLEEGVEYEFVRSVENIVGIGKPSKVSCEYVARDPCDPPGR	15948	Db	16969	SILWTKVNKTIIHDTQFKAQNLKEGIEYEFVRSVENIVGVKASKNSECYVARDPCDPPG	17028
QY	290	-----	289	QY	300	-----	299
Db	15949	EAIIVTRNSVTLQWKPTVDGSKITGYIVEKKELPEGRWMAKSFNIIIDTHEVETGLVE	16008	Db	17029	TPEPIMVKRNEITLQWTKPEVYDGGSMITGYIVEKRDLPGRWMAKSFNIIIVTQFTVSG	17088
QY	290	-----	289	QY	300	-----	299
Db	16009	DHRYEFRVIARNAAGVSEPSSESTGAITARDEVPDPRISMDPKYKDTIVVHAGESFKVDA	16068	Db	17089	TEDQRYEFRVIAKNAAGAIKSPDSGTPIAKDEVELPRISMDPKFRDTIVVNAGETFRL	17148
QY	290	-----	289	QY	300	-----	299
Db	16069	DIYGKPIPTIOWIKGQELSNARTARLEIKSTDPAFSLVKDAVRVDSGNYILKAKWAGER	16128	Db	17149	EADVHGKPLPTIWLKRGDXEIBESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAG	17208
QY	290	-----	289	QY	300	-----	299
Db	16129	SVTNVKVLDRPGPEPVVISGVTAETKCTLAWKPLQDGGSDIINYIVERRETSRLVMT	16188	Db	17209	SKSPFVNVKVLDRPGPEPVQVGTSEKSLTWSPPLODGGSDISHYVVEKRETSRLA	17268
QY	290	-----	289	QY	300	-----	299
Db	16189	VVDANVQTLSCVKTKLEGNEYTFRIMAVNKNYGVGBPLESEPVVAKNPVVPDAPKAPEV	16248	Db	17269	WTVVASEVNTSLKVTKLLEGNEYPRIMAVNKNYGVGBPLESAPVLMKNPFVLPGPCKSL	17328
QY	290	-----	289	QY	300	-----	303
Db	16249	TTVTKDSMIVWERPASDGGSEILGYVLEKRGKEGIRWTRCHKLLIGELRLRTVGLIENH	16308	Db	17329	EVTNIAKDSMTVCWNRPSDGGSEIIGYIVEKDRSGIRWIKCNKRITDLRLRVTGLTE	17388
QY	290	-----	289	QY	304	-----	311
Db	16309	DYEFVRSNAAGLSBSPSPAYQKACDPIYKPGPNPKVIDITRSSVFLSKSPIYDG	16368	Db	17389	DHEYEFVRSNAAGVGEPSATVYKACDPVKPGPPTNAHIVDTTKNSITLAWKPIY	17448
QY	290	-----	289	QY	312	-----	311
Db	16369	GCEIQGYIVEKDVNVGEWTMCTPPTGINKNTIEVEKLLKHEYNFRICAINKAGVGEHA	16428	Db	17449	DGSEILGYVVEICKADEBEWQIVPTQGLRVTFRPEISKLEHQBKIRVCALNKVGLGE	17508
QY	290	-----	289	QY	312	-----	311
Db	16429	DVPGPIIVEKLEAPDIDLLELRKIINIRAGGSLRLFVPIKGRPTPEVKWGVGDGEIRD	16488	Db	17509	ATSVPGTVKPEKLEAPDLDELRLKGIWVAGGSARIHIPKGRPMPEITWSRBEGBF	17568
QY	290	-----	289	QY	312	-----	311
Db	16489	AAIIDVTSSFTSLVDNVNRYDSKVTLTLENSSGTKSAFVTVRVLDTPSPVNLKVTEI	16548	Db	17569	TDKVQIEKGVNTQSLSIDNCDNDAGKYLKLENSSGSKSAFVTVKVLDTPGPPQNLAVK	17628
QY	290	-----	289	QY	312	-----	311
Db	16549	TKDSVITWEPPLLDGGSKIKNYIVEKREATRKSAAVVTNCHNSKWKIDQLQEGCSYF	16608	Db	17629	EVRKDSAFVLEWPPPIIDGGAKVKNYVIDKRESTRKAYANVSSKCKSTSFKNVLETSYIY	17688
QY	290	-----	289	QY	312	-----	311
Db	16609	RVTAENEYGIGLPAQTADPIKVAEVQPPGKITVDVTRNSVLSWTKPEHDGGSKIIQY	16668	Db	17689	YFRVMAENBEFGVGPVETVDVAKAAEPSPGKVTLTDSYQTSASLWKEKPEHDGGSRLV	17748
QY	290	-----	293	QY	312	-----	311
Db	16669	IVEMQAKHSEKSEKARVKSQAVITNLTOGEBYLFVAVVAVNEKGRSDPRSLAVPIVAKD	16728	Db	17749	GYVEMQPKGTEKWSIVAESKVCNAVVTGLSSGQEQYQFRVKAINEKSGSDPRVLGVFVIA	17808
QY	294	-----	293	QY	312	-----	311
Db	16729	LVIEPDVKPAFFSYVQGDQKIEVPIISGRPKPTITWTKOGULPKQTTRINVTDSLDT	16788	Db	17809	KOLITQPSLKLFPNTYSIOAGSBLKIEIPVIGRPRPNISWVKDGEPLKQTRVNVSETAT	17868
QY	294	-----	299	QY	312	-----	311
Db	16789	TLSIKETHKDDGGQYGITVANVVGQKTASIEIVTLKDPDPKGPVDFDVSAESITLSWN	16848	Db	17869	STVLHIKEGNKDDFGKYTVTATNSAGTATENLSVILEKPGPVGPVDFDEVSADFVWIS	17928
QY	300	-----	299	QY	312	-----	311
				Db	17929	WBPPAYTGCGCOISNYIVEKRDTTTTTWHMVSATVARTTIKTKLTGTGYQFRIIPAENRY	17988

QY	312	-----	311	QY	327	-----	326
Db	17989	GKSAPLDSKAVIQYPFKEPPGPTFTSISKDQMLVOWHEFPVNDGGTKIIGYHLEQXE	18048	Db	19069	RYGKSSYSESSAVVAEYFSPGPGTQKVHATKSTMLVTWQVPVNDGSRVIGYHLEY	19128
QY	312	-----	311	QY	327	-----	326
Db	18049	KNSILWVKNKTIQDTKTKTGLDSGLEVBKVSANENVIGIKPSKVSECFVARDPCDP	18108	Db	19129	KERSILWSKANKILIADTQVKVSGLDGLMYRVRVAENIAGIGKSKSEFVPARDPC	19188
QY	312	-----	319	QY	327	-----	326
Db	18109	PGRPEAIVTRNNVTLKMKPAYDGGKITGYIVEKKOLPDGRWMMKASFNVLETFVTS	18168	Db	19189	DPPQPEVTNITRKSLSLWSPHYDGGAKITGYIVERRELDPGRWLKCNNTNIOETYPE	19248
QY	320	-----	319	QY	327	-----	326
Db	18169	GLVEDORYEFRVJARNAGNFSEPSGAIHTARDEIDAPNASLDPKYKDVIVVHAGTF	18228	Db	19249	VTELTEDQRYEFRVFARNADSVSBPSESTGPIIVKDDVEPPRVMMDVKFRDVIIVVKAGE	19308
QY	320	-----	319	QY	327	-----	326
Db	18229	VLEADIRGKPIPDVWWSKDGKELEETAARMEIKSTIQKTTLVVKDCIRTDGGQYILKLSN	18288	Db	19309	VLKINADIAGRPLPVISWAKDGIIEERARTEIISTDNHTLLTVKDCIRRTDQGVVLTK	19368
QY	320	-----	319	QY	327	-----	326
Db	18289	VGGTKSIPITVKLDRPGSPGPKVTGVTAEKCYLAWNPLODGGANISHYIIEKRETS	18348	Db	19369	NVAGTRSVAVNCKVLDKPGPPAGPLEINGLTAEKCSLSWGRPOEDGGADIDYVHKKRET	19428
QY	320	-----	322	QY	327	-----	326
Db	18349	RLSWTQVSTEVQALNYKVTKLLPCNEYIFRVMAVNVKYGIGEPLESPVTACNPYKPPGP	18408	Db	19429	SHLAWTICEGELQMTSCKVTLLKNGEYIFRVTVGNKYGVGEPLSEVAIKALDPFTVPSP	19488
QY	323	-----	326	QY	327	-----	326
Db	18409	STPEVSALTCDMSVVTWARPVDDGGTEIEGYILEKRDKEGVRWTKCNKKTLDLRLRVTG	18468	Db	19489	PTSLEITSVTKESMTLCSWRPESDGGSEISGYIERREKNSLRWVRVNVKFPVYDLRVKST	19548
QY	327	-----	326	QY	327	-----	326
Db	18469	LTEGHSYEFRAAENAAAGVEPSEPVFRACDALYPPGPPSNPKVTDTSRSSVSLAWSK	18528	Db	19549	GLRECEYEYRVYAENAAAGLSLPSSETSLIARAEDPVFLPSPSPKPIVDSGKTITIAWV	19608
QY	327	-----	326	QY	327	-----	326
Db	18529	PIYDGGAPVGYVEVEKEAADEWTTCTPTGLQKQFTVTKLKENYENFRICAINSEG	18588	Db	19609	KPLPDGGAPITGYTVVEYKSDDTWKTSIQSLRGTEYITISGLTTGAEYVFRVKS NVKVA	19668
QY	327	-----	326	QY	327	-----	326
Db	18589	VGEPATLPGSWAQERIEPPEIELDADLRKVVLRASATLRLFTVTKRPEPEVKWEKAE	18648	Db	19669	SDPSDSDPQIAKEREEREPFLDIDSEMRKTLIVKAGASFTMTVPRGRPVPNVLSKPT	19728
QY	327	-----	326	QY	327	DIAT	330
Db	18649	GILTDRAQIEVTSFTMLVIDNVTRFDSGRYNLTLNNSGSKTAFVNVRLDPSAPVNL	18708	Db	19729	DLRTRAYVDTTDSRTSLTIENANRNDSGKYTLTIQNVLSAASLTLVWKVLDTPGPPTNIT	19788
QY	327	-----	326	QY	331	-----	330
Db	18709	TIREVKDSVTLSEPPIDGGAKITNYIVEKETTRKATITNNCTKTTFRLENLQEG	18768	Db	19789	VQDVTKE SAVLSWDVPENDGGAPVKNYHIEKREASKAWSVVTNNCNRLSYKVTLNQEGA	19848
QY	327	-----	326	QY	331	-----	330
Db	18769	CSYVFRVLASNEYIGLPAETTEPVKVSPEPLPPGRVTLVDVTRNTATIKWEKPESDGGS	18828	Db	19849	IYVFRVSGENEPGVGIPAEKTEGVKITEKPSPEKLGVTISISKDSVSLTWLKEPDGGR	19908
QY	327	-----	326	QY	331	-----	332
Db	18829	KITGYVVMQTKSEKSWSTCTQVKTLBATISGLTAGEEYVFRVAANVEKGRSDPRQLGVP	18888	Db	19909	IVHYVVEALEKQKNWKCVAKSTHHVVVSGLSENSEYFFRVFAENQAGLSDPRELLLPV	19968
QY	327	-----	326	QY	333	-----	332
Db	18889	VIARDIEKPSVELPFHTFNVKAREQLKIDVPKGRPQATVNVNRKDGQTLKETTRNVSS	18948	Db	19969	LIKEQLEPPIDMKNFPSSHVVVRAGSNLKVDIPISGKPLPKVTLSRDGVPLKATWRFNT	20028
QY	327	-----	326	QY	333	-----	332
Db	18949	SKTWTLSIKEASKEDVGTIELCVSNSAGSITVPITIIIVLDRPCCPPPIRIDEVSCDSIT	19008	Db	20029	EITAENLTINKESVTADAGRVEITAAANSSGTTKAFINIVLDRPCCPPTGPVVVISDITEE	20088
QY	327	-----	326	QY	333	-----	332
Db	19009	ISWNPPEYDGGCQISNIVEKETTSTTWHIVSQAVARTSIKIVRLTTGSEYQFRVCAEN	19068	Db	20089	SVTLKWEPPKYDGGSOVTNYILLKRETSTAVWTEVSATVARTMVKMKLTGTEEYOFRIK	20148
				QY	333	-----	335

Db	20149	AEHRFGISDHISACVTVKL		21229	EYQFRVRAENRYGV	SOPLVSSII	VAKHQFRI	PGPPGKPI	YNYNTSDGMSLTWDAPYDGG	21288
QY	336	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20209	LEMKDRNSILWQANKLVIR	THFKVTTISAGLI	YEFRVYAENAGVKPSHP	SEPLAI	20268				
QY	336	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20269	DACEPPNRVRI	TDISKNSVLSWQPAFDG	SGKITGYI	VERRDL	PDGRWTKASFTNV	TET	20328		
QY	336	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20329	QFTISGLTQNSQY	FRVAFARNAGVSNP	SVGPITC	IDSYGP	VIDLPLEY	TEVVKYR	20388		
QY	340	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20389	AGTSVKLRAGISG	KPAPTIEWYKDK	ELQTNALVC	VENTD	LASILIKD	ADRLNSCYEL	20448		
QY	340	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20449	KLRNAMESATIR	QILDKPGPGGPI	EBKTVTA	EKITLLMR	PPADGGAKITHI	VEK	20508		
QY	340	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20509	RETSRVVMSV	SEHLEECITTTKII	KGNEYIP	RVRAVNKY	GIGEP	LES	20568		
QY	340	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20569	PGPPGIP	EVTKITNSMTV	WSRPIADG	SDISGY	FLEKRD	KSLG	20628		
QY	340	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20629	KVTGLTENS	DYQVRVCA	NAAGQGP	SEPEFY	KAADPID	PGPPAKIRI	20688		
QY	348	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20689	GMSKPYDGS	AVTGYVVEI	ROQEEB	WTVT	STKGEV	RTTEYV	20748		
QY	348	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20749	NCAGQGEPI	EMNEPVQAK	ILEAPE	IDLVAL	RTSVIA	KAGEDV	20808		
QY	348	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20809	RKDEKNL	SDARYSI	ENTDSS	LLTI	PQVTR	NDTGK	20868		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20869	PAAQKQV	KHVS	RGTV	TLMDP	PLIDG	GSPI	20928		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20929	LIDLSEK	TPFRVLA	ENEIG	EGPC	ETTEP	VPVAA	20988		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	20989	PDFDGS	VITEYV	VERK	GEQ	TSW	HAGIS	21048		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21049	PVTIGP	ITVKELI	ITPE	VDLSD	IPGAQ	VTVR	21108		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21109	ESEFV	RFSKTEN	KITLSI	KNKKE	HGKVT	VILD	21168		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21169	FDEIKAD	SVILS	WDV	PNED	GGGEI	TCYSI	21228		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21229	EYQFRVRAENRYGV	SOPLVSSII	VAKHQFRI	PGPPGKPI	YNYNTSDGMSLTWDAPYDGG	21288			
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21289	SEVTGFHVEK	KERN	SILWQ	KVNTSPIS	GREY	RATGL	21348		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21349	SKFTLAV	SPVDP	PGTPDY	IDVT	TRETTIL	KWNP	21408		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21409	TDVSECQ	YTVTGL	SPGDR	YEPRII	ARNAVGT	ISPP	21468		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21469	DGLI	IKSG	BSLRI	KALVQ	GRPV	RTWFKD	21528		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	351
Db	21529	RGYTV	AEAK	NSAKAEI	KVKVQ	DTPTG	KVGP	21588		
QY	352	-----	-----	352	-----	-----	-----	-----	-----	356
Db	21589	HYII	EKRETS	RLAW	ALIEDK	CEAQ	SYTAI	21648		
QY	357	-----	-----	357	-----	-----	-----	-----	-----	356
Db	21649	QIQY	TVDP	APGI	PEPS	NITGNS	ITLT	21708		
QY	357	-----	-----	357	-----	-----	-----	-----	-----	356
Db	21709	PISE	TRFK	VTGL	TEG	NEYEF	HVMA	21768		
QY	357	-----	-----	357	-----	-----	-----	-----	-----	356
Db	21769	SKT	TVS	LE	SKVP	FDG	MEIIGY	21828		
QY	357	-----	-----	357	-----	-----	-----	-----	-----	363
Db	21829	FRV	SAI	AG	KD	SC	EV	21888		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	21889	TPT	AV	MSK	PD	SN	LS	21948		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	21949	LDT	PG	PG	PI	TF	KD	22008		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	22009	IF	KV	ND	LA	EG	VP	22068		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	22069	WL	KPD	H	DG	SG	RI	22128		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	22129	YSE	PRE	AF	SS	VI	IK	22188		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	22189	MR	L	K	ET	D	R	22248		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	363
Db	22249	GP	IE	V	SS	VA	ES	22308		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	370
Db	22309	TK	Y	E	S	F	R	22368		
QY	364	-----	-----	364	-----	-----	-----	-----	-----	370

QY	371	-----	370	QY	382	-----	381
Db	22369	HDGSKIIGYWEKBERNTILWVKENKVCLECNKYKVTGLVEGLEYQFRTYALNAAGVSK	22428	Db	23449	DGMLTWYPEDDGGSOVTGYIVERKEVRADRWVRNKKVPTWTRYSTGLTEGLEYEHR	23508
QY	371	-----	370	QY	382	-----	381
Db	22429	ASEASRPIMAQNVPDAPGRPEVTDVTRSTVSLIWSAPAYDGGSKVVGIIERKPVSEVD	22488	Db	23509	VTAINARGSGKPSRPSKPIVAMDPDAPPKQPNQPRVTDTRTSVSLAWSVPDEGGSKVT	23568
QY	371	-----	370	QY	382	-----	381
Db	22489	GRWLKCNVTIVSDNFTVTALSEGDTYEPRVLAKNAAGVISKGSBSTGPTCRDEYAPPK	22548	Db	23569	GYLIEMQVDQHEWTKCNTTPTKIREYTLTHLPQAEYRFRVLACNAGGPEPAEVPCTV	23628
QY	371	-----	370	QY	382	-----	389
Db	22549	AELDARLHGLVTIRAGSOLVDAAVGGKPEPKIIWTKGDKELDLCCKVSLQYTKRATA	22608	Db	23629	KVTMLEYPDYELDERYQEGIFVRQGGVIRLTIPKIKPPFPCKWTKGQDISKRAMIAT	23688
QY	371	-----	370	QY	390	-----	392
Db	22609	VIKFCDRSDGKTLTVKNASGTVKAVSMVKVLDSPGCGKLTVRVTOEKCTLAWSLPQ	22668	Db	23689	SETHTELVIKEADRGDSGTVDLVLENKCGKANVYIKURVIGSPNSPBGPLEYDDIQVRSV	23748
QY	371	-----	370	QY	393	-----	392
Db	22669	EDGGBIITHYVERRETSLRNWVIVEGECPTLSYVVTRLIKNNEXIFRVRVKNYGPVGP	22728	Db	23749	RVSMRPPADGGADILGYILERREVPKAAWYTIIDSRVGTSLVVKLKENVEYHFRVSAE	23808
QY	371	-----	370	QY	393	-----	392
Db	22729	VESEPIVARNSTIPSPGPIPEEVGTGKEHIIQWTKPESDGNELSNLYLVDKREKESLR	22788	Db	23809	NQFGISKPLKSEEPVTPKTLNPPPEPSNPPEVLDTKSSVLSWSRPKDDGGSRTVGY	23868
QY	371	-----	370	QY	406	-----	406
Db	22789	WTRVNDYVYDTRLKVTSILMECDYQFRVAVNAAGNSEPSSNFISCRPSYTPGPP	22848	Db	23869	IERTSTDKVRHNKQITTTMYTVTLVDAEYQFRIIAQNDVGLSETSPASEPVVCK	23928
QY	371	-----	370	QY	407	-----	409
Db	22849	SAPRVVDTTKHSISLAWTKPMYDGTDIVGYVLEMQEKDQDQWYVHTNATIRNTEFTVP	22908	Db	23929	DPFDKPSQGELEILLSISKDSVTLQWEKPECDGKBEILGYWVEYRQSGSAWKSKNERI	23988
QY	371	-----	370	QY	410	-----	409
Db	22909	DLKMGQKISPRVAANVKGMSSEYSEIAEIEPVERIEIPDLLELADDLKKTVIRAGASLR	22968	Db	23989	KDKQFTIGLLEATEYEFVRFAENETGLSRPRRTAMSIKTLTSGEAPGIRKEMKDVTK	24048
QY	371	-----	370	QY	410	-----	409
Db	22969	LMVSVSGRPPPVITWSKQIDLASRAIIDTTESYSLLIVDKVNRVDAGKYTIEAENQSGK	23028	Db	24049	LGEAAQLSCQIVGRPLPDIKWYRFGKELIQSRKYKMSDGRHTLTVMTEQEGEVYTC	24108
QY	371	-----	370	QY	410	-----	409
Db	23029	KSATVLVKVYDTPGPCPSVKVKSRSVITITWEIPTIDGAPINNYIVEKREAAARPK	23088	Db	24109	IATNEVGEVETSSKLLQATPFQHPGYPLKEKYGAVGSTLRLHVMYIGRPVPAMTWFG	24168
QY	371	-----	374	QY	410	-----	409
Db	23089	TVTTKCSKTLRISGLVEGTMHYFRVLPENIYIGIBPCETSDAVLSEVPLVPAKLEWVD	23148	Db	24169	QKLLQNSENITIENTEHTYTHLVKQVQRKTHAGKYQVLSNVFGTVDAILDVEIQDKPK	24228
QY	375	-----	374	QY	410	-----	409
Db	23149	VTKSTVTLAWKPLYDGSRLTGYLBACKAGTERMMKVTLKPTVLEHTVTSLSNEGEQY	23208	Db	24229	PTGPIVIEALLKNSAVISWKPADDGGSWITNYVVEKCEKAGEBQLVSSAISVTTCRI	24288
QY	375	-----	381	QY	410	-----	417
Db	23209	LFRIRANEKGVSEPRETVTAATVQDLRVLPDLSLTPQKTHVPAGRVELVPIAGR	23268	Db	24289	VNLTENAGYFRVSAQNTFGISDPLEVSSVIIKSPFEKPGAPGKPTTAVTKDSCVWAW	24348
QY	382	-----	381	QY	418	-----	417
Db	23269	PPPAASWFFAGSKLRESERVETHTKVAKLTIRETTIRDTGEVTLBKNVTGTTSETIK	23328	Db	24349	KPPASDGGAKIRNYILEKREKQNKWISVTTEIRETVFSVKNLIEGLEYEPRVKCNLG	24408
QY	382	-----	381	QY	418	-----	417
Db	23329	VIIIDKPGPTGPIKIDIBDATSITISWEPPELDGGAPLSGVYVQORDAHRPGWLPVSES	23388	Db	24409	GESEWSEISEPITPKSDVPDPIQAPHKEELRNLRVYQSNATILVCKVTHGHPKIVKWYRQ	24468
QY	382	-----	381	QY	418	-----	417
Db	23389	VTRSTFKFTRLTGEGYVFRVAATNRFGISYLOQSEVIECRSSIRIPGPPETLQIPDVS	23448	Db	24469	KEIIADGLKYIOEFKGYHQLIIASVTDDATVQVVRATNOGGSVSGTASLEVEVPKI	24528
				QY	418	-----	417

Db	24529	HLPKTLGMAVHALRGEVVISIKIPFSGKPDVPVITWKGQDLIDNNGHYQVIVTRSFSL	24588	Db	25609	KRTEERLLEBELELGFSPSPSRPPHPELSSRLSYSSPQAHVKVEETRNFRSYTHI	25668
QY	418	-----	417	QY	437	-----	436
Db	24589	VFPNGVERKDAGFYVYVCAKRFIDQKTVELDVADVPDPGRGVKVS DASRDSVNLWTWEP	24648	Db	25669	PTKABASTYAELELREHQAAYRQPKQRORIMAREDEELLRPVTTTQHLSEYKSELDFM	25728
QY	418	-----	417	QY	437	-----	443
Db	24649	ASDGSKITNYIVEKCAATTAERLVRQARETRYVINLFGKTSYQFRVIAENKFLGSKP	24708	Db	25729	SKEEKSRRKRRQREVTEITEIEEYEYISKHAQRESSASRLLRRRSLSPYIELMRP	25788
QY	418	-----	417	QY	444	-----	445
Db	24709	SEPSPTITKEDKTRAMNVDBEVEDRETVSMTKASHSTKELYEKYMIADLGRBFGIV	24768	Db	25789	VSELIRSRQPAABEYEDDTERRSPTPERTRPSRSPVSSERSLSRFSRGPDI FSRYES	25848
QY	418	-----	417	QY	446	-----	445
Db	24769	HRCVETSSKTYMAKFVKVGTDOVLVKEISILNIAHRNHLHESFESMEELWIFE	24828	Db	25849	MKAALKTKTSEKYEVLQQPPTLDHAPRITLRMRSHRVPQCQNTFRILNVQSKPTAEV	25908
QY	418	-----	417	QY	446	-----	445
Db	24829	FISGLDIFERINTSAFELNEREIVSVHVCEALQFLSHNIGHDIRPENIIYQRRSS	24888	Db	25909	KWYHNGVELQESSKIHNTSGVLTLLEILDCHTDDSGTYRACVCTNYKGEASDYATLDVTG	25968
QY	418	-----	418	QY	446	-----	445
Db	24889	TIKIEFGOARQLKPGDNFRLFTAPEYYAPEVHVDVSTADMMWSLGLTVYVLLSGIN	24948	Db	25969	GDYTYVASORRDEEVPRSVFPBELTRTEYAVPSPFKTSEMEASSSVREVKSQMTETRESL	26028
QY	419	-----	425	QY	446	-----	445
Db	24949	PFLAETNQIENIMNABYTFDEAFKEISIEAMDFVDRLLVKEKSRMTASEALQHPWL	25008	Db	26029	SSYEHSASAEMKSAALEKSEKSTTRKIKTTLAARILTKPRSMTVYGESARFSCDTD	26088
QY	426	-----	425	QY	446	-----	445
Db	25009	KQKIERVSTKIVIRTLKRRYHHTLKKDLNMVVSAAISCGGAIRSQKGVSAKVAVASI	25068	Db	26089	GBPVPVTWLRKGQVLSTARSQVTTTKYKSTFEISSVQASDEGNYSVVVENSEGQAE	26148
QY	426	-----	433	QY	446	-----	452
Db	25069	EIGPVSQIMHVGEGGHVKYCKIENYDQSTQVTFVGRQLENSBYEITYEDGVAI	25128	Db	26149	FTLTQKARVTEKAVTSPPRVKSPEPRVKSPEAVKSPKRVKSPSPHPKAVSTETKPTP	26208
QY	434	-----	433	QY	453	-----	452
Db	25129	LYVKDITKLDGTYRCVNDYGEDSSYAELELVKGVREVDYCYCRTMKIKERTDTMRL	25188	Db	26209	REKVQHLPVSAAPPKITQPLKAEASKEIAKLTCTVWESSVLRAKEVTVWKDQKCLKENGHFQ	26268
QY	434	-----	433	QY	453	-----	458
Db	25189	LERPPEFTLPLYNKTA YGVENRFGVTITVHPBPHVTWYKSGQIKPGDNDKKYTPESDK	25248	Db	26269	PHYSADGTYELKINNLTESDQGEYVCEISGEGGTSKTNLQFMGQAFKSIHEKVSKISETK	26328
QY	434	-----	433	QY	459	-----	461
Db	25249	GLYQLTINSVTTDDAEYTVVARNKYGEDSCAKLTVTLLHPPTDSTLRPMPKRLLANAB	25308	Db	26329	KSDQKTTESTVTRKTEPKAPBPISSKPVI VTGLQDTTVSSDSVAKPAVKATGPRPTAIW	26388
QY	434	-----	433	QY	462	-----	461
Db	25309	CQEGQVCFEIRVSGIPPPTLKWEKDGQPLSGPNIEIIEHGLDYALHIRTLPEDTGY	25368	Db	26389	TKDGKAITOGGKYKLSKEDKGGFFLEIHKHTDTS DGLYTCVKNAGSVSSSCKLTIKAIK	26448
QY	434	-----	433	QY	462	-----	461
Db	25369	YRVTAINTAGSTSCQAHLQVERLRKYKQEFKSEHERHVQKQIDKTLRMAELTSGTESV	25428	Db	26449	DTEAQKVSTQKTSEITPQKAVVQBEBISOKALRSEBIKMSAKSOBKALKEBASKVLIS	26508
QY	434	-----	433	QY	462	-----	461
Db	25429	PLTVAKEALREAAVLYKPAVSTKTVKGBFLBIEBKBERKLRMPYDVPEPRKYQTITI	25488	Db	26509	BEVYKSAATSLBKSIVHBEITKTSQASEVRTHAEIKAFSTQMSINEGORLVJKANIAGA	26568
QY	434	-----	433	QY	462	-----	462
Db	25489	EEDQRIKQFVMSDMKWKYKIRDOYEMPGKLD RVQKRPKRIRLSRWEQFYVWPLPRIITD	25548	Db	26569	TDVKVNLNGVELTNSREYRYGVSGSDQTLTIKQASHRDESGILTCISKTKEGIVKCYQDILT	26628
QY	434	-----	433	QY	463	-----	465
Db	25549	QYRPKWRI PKLSQDDLEIVRPARBRTSPDPYDFYRPRRSLGDISDEHLLLPIDDYLAM	25608	Db	26629	LSKELSDAPAFISQPRSONINEQNVLF TCEISGEPSPIENWFK	26672
QY	434	-----	436				

Thu Jan 6 14:34:13 2005

ID XX ABM67171 standard; protein; 16368 AA.
AC XX ABM67171;
XX XX
DT 20-NOV-2003 (first entry)
DE XX
DE XX Photorhabdus luminescens protein sequence #268.
XX XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX XX
XX Photorhabdus luminescens.
OS XX
XX WO200294867-A2.
XX XX
XX 28-NOV-2002.
XX XX
XX 07-FEB-2002; 2002WO-IB003040.
XX XX
XX 07-FEB-2001; 2001FR-00001659.
XX XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX WPI; 2003-148459/14.
XX XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX XX
XX Claim 2; SEQ ID NO 268; 1205pp; French.
XX XX
XX The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX XX
SQ Sequence 16368 AA;
Query Match 23.7%; Score 581; DB 6; Length 16368;
Best Local Similarity 1.6%; Pred. No. 0.00032;
Matches 246; Conservative 88; Mismatches 131; Indels 14861; Gaps 84;
QY 1 MGRKE-----5
DB 996 LGENDQOVKIRGRIEPIEIRLMEHPAVSEALVIALNDGQKELVAVYMAEADDQLVN 1055
QY 6 -----5
DB 1056 SLRAYLSAILPDYMWPTAFVRLDTFPLTPNGKLDRRALPAPGEAFARQAYAAPQGETES 1115
QY 6 -----MMVRDV-----11

Db 1116 LLSAVMGELIGIEQVSRHDSFFALGCHSLAVRMVERLNRNAGLTLAVRDLFOSPVLSEFA 1175
QY 12 -----PKMFVLISIS-----21
Db 1176 QTVGSOAVIVPANIMTPAITALTPEMLPLIDLTSQSEIDHIVAQVPGGVANIQDIYGLSP 1235
QY 22 -----FLLV-----26
Db 1236 LQDGLPHLLANKGDPYLFVSKWVFADRALDLSYLVAVQVVRNHDILRTAFVWGLSA 1295
QY 27 -----26
Db 1296 PVQVWRQAQLSVTELTDPVDPGVRDQLAQYFDLHHRLDLNQAPLLHFGVAQEEDGRW 1355
QY 27 -----26
Db 1356 SVLQLLHLIGDHTTLEVMNGEIQAYLAGQEGNLPVPIPPRLNVAQARLGVSQAEHTRFF 1415
QY 27 -----26
Db 1416 TEMLAEDVQPTLPFGLMEVHRDGSQVTSYRMLPATLNDRLRSQAORLGVSLATLCHLAW 1475
QY 27 -----FIN-----29
Db 1476 AQLSHTSGQEKVVGTVLFGRMVQGGADSGMGLFINTLPLRLDINDASVDQSVQAVHT 1535
QY 30 -----29
Db 1536 RLAEILDHEHMSLALAQCRSRVPGEIPLFSALLNYRHNDQTFPENKEVPGIEFIGAERT 1595
QY 30 -----CKWMSKAL-----37
Db 1596 NYPFVLSVEDGGSTGLGTAQVQVPPFESERVCGYMQQALLESALAEALQRPETPTVTLNLP 1655
QY 38 -----37
Db 1656 AAEKKLLLETWNATATETSYDPHSCIHRLFEQWAKKSPDATAVLYVEEQLISYANLNACAN 1715
QY 38 -----37
Db 1716 RLAHQIALGVTPDQORVAVCVSRPAMVAGVAVLAKAGAVPLDDPAPSPARLAHLTDA 1775
QY 38 -----37
Db 1776 APANVLADKAGKVALGEBEALTGLTVLDNLPNQVDSNPQISGLTSRHLAYVIVTSGSTG 1835
QY 38 -----37
Db 1836 VPKGWMIEHRTNVNFLYWAQAFETEIREVLFTSMNFDLSIFECPVLSOGATIHLE 1895
QY 38 -----37
Db 1896 DALSLMQHALPVTLINSVPSAMKPLLQALMASVHTVNLAGEPLKGLIEQIFEKTOIQ 1955
QY 38 -----37
Db 1956 RLCNLYGPSETTYSANWLFQGRDRIIESIGRIPIANTRLYLDDENGQPVPLGVGVEIYIG 2015
QY 38 -----YNRP-----41
Db 2016 GAGVARGYFNRPDLTAERFLIDPFDSDVPDARMYRTGDLARYPDGNLEFLGRNDQOVKIR 2075
QY 42 -----41
Db 2076 GFRIEPEGEARLMEHPAVSEAVILALDDDDQDKRLVAVYVAEADVLINRLHAHLASVLP 2135
QY 42 -----WRGLV-----46
Db 2136 DYMPAFAVCLDAPPLTPNGKLDRRALPAPGEEDFARQIYAAPSGEMETTLAIAIMRELIG 2195
QY 47 -----LSKIG-----51

Db	2196	VERISRYDNFFALGCHSLLAVRMNRI	AALGIELPLSTLFTFP	SLJAPAEVMTQFDESG	2255	QY	64	-----	63
QY	52	-----	-----	-----	51	Db	3336	DHIEQPGGVANIQDIYALS	PLQDGIPLFHLANKGDPYLLVSQMI
Db	2256	AALPAIVPLSREDHQPLSPAQQRLWFLAQLNGVSEYTHIPLALRLRGRLDITAMQQALD	2315	QY	64	-----	63		
QY	52	-----	-----	-----	58	Db	3396	VOOVNRHILRTAFWQGLSAPVQVVRQAQLSVTELTD	PVDGVPVRQLAQHPDPRH
Db	2316	TLFARHEALRSVFAVDGQPVVELLPAASSLLMKYDLRDTDPDQOLAFLSAQEASAPF	2375	QY	64	-----	63		
QY	59	-----	-----	-----	58	Db	3456	RLDLNAQPLLHFAIAQEDDGRWSVLQLLHHLIGDHTTLEVMNGEVQAYLAGQESLPAPA	3515
Db	2376	DLARGPLIRSSVLQADDEHVFLITQHHIVSDGWSLGLKSELEALYSALHNLQPPPLP	2435	QY	64	-----	63		
QY	59	-----	-----	-----	58	Db	3516	PFRLVAQARLGVSAEHTRFTFDM	LAEVEPTLPFGLTEVYRDGSQVTEFHRMLPATLN
Db	2436	LTIQPDYAAQOROWLSAERIQSOSDYWRARLVDA	PVLLDLPTDRPPQPSFAGALLPV	2495	QY	64	-----	72	
QY	59	-----	-----	-----	58	Db	3576	DRLSQARLDSVLTATLCHLAWAQLVSR	TSGREVVFGTVLFRMQAGOGADSGMGLFIN
Db	2496	NLDAELTQSLKRLSEQGVTLFWTLAAWTA	VLRSLSGQEDLVVGTSPSAGRSRQEVESLI	2555	QY	73	-----	77	
QY	59	-----	-----	-----	58	Db	3636	TLPLRLDMDETSVQDSVRLAHTRLAGLLE	HEHASLALAQRCSGVLGEVPLFSALLNYRHS
Db	2556	GFVNTLALRMDLSGAPNUTELLARVQ	TALAAQEHQDLPFEQVVEIVQPPRLAHTPLP	2615	QY	78	KQPV	-----	81
QY	59	-----	-----	-----	58	Db	3696	TQPVTTDAIISGIEFSEQERTNYPFVLS	VEDFGHALGLTQIVQIPDPERICGYMQAL
Db	2616	QVIFAWQNNENIEWRLPGLAVSLADQPIDIAK	FDLSLSLSEVDGRVGVLYNATAPDQA	2675	QY	82	-----	88	
QY	59	-----	-----	-----	63	Db	3756	ESLVEALQAPETPIRALNILPAEKRLL	LETWNATQAPYDQCCVHQLFEOQAGNTPDA
Db	2676	TIERQVGYLTVLRGMATNPQPVGKIDILTV	TERKLLLETWNATKTSYPDHGCIHRLFE	2735	QY	89	-----	88	
QY	64	-----	-----	-----	63	Db	3816	IALVYGEHLTSAELNARANWLAHMLI	GQIQDORVAVQLERSIELVVAQLAILKAGAV
Db	2736	QQAECTPDALVYKYLISYAELNARANLAH	QLITLGVAPQORVAICVASSPARIVGL	2795	QY	89	-----	95	
QY	64	-----	-----	-----	63	Db	3876	YVPIDRPVDERKHWLISDCAAKLLITDI	PVLADKIGAISGEDYENPDLPRISTELAYI
Db	2796	LAVLKAGAVVPLDPAYPGERLHI	LTDAAPIVLADNTGRAALGKVLAAALT	2855	QY	96	L	-----	99
QY	64	-----	-----	-----	63	Db	3936	MYTSGSTGPKGVLPHRAVARLVINNG	VAEIPGDDRVAFGANPSPDASTFEVWAPLLNG
Db	2856	PQPDNSNPQVPLTPHLLAYVIYTS	GTGIPKGVWVEHCGLVNLVQEKIVQFDIHPGSRM	2915	QY	100	-----	109	
QY	64	-----	-----	-----	63	Db	3996	GTLLVIDHATVLTPEPALALQTYRITV	LMVLSVGLFNRLVAELSPLVLPQLKILIVGSDVL
Db	2916	LQFASFGDAGVWETMMALCSGATLAI	PADTVRQEPRLWHYLEBRAITHACLTPALLRE	2975	QY	110	-----	109	
QY	64	-----	-----	-----	63	Db	4056	DPHVMAQVLRDNPQQLLSAYGSP	EGTFTTTCITALPSSVARIPIGRPIANARIYLLD
Db	2976	GTDLPEMTIRPTLILGGEAPSTLLQAL	CERATVFNAYGTEITVCAATWRCPSDYTEGV	3035	QY	110	-----	109	
QY	64	-----	-----	-----	63	Db	4116	TYGQPVPLGAIGEIVYGGDGVACGY	LNRPDQTAERFLTPDPSQPDARMYRTGDLARYLP
Db	3036	IAIGRPANTQIYLLNTDQPVPLGAV	GELYIGGIWARGYLNRPDLTAERFLADPFSDK	3095	QY	110	-----	109	
QY	64	-----	-----	-----	63	Db	4176	DGNLEFLGRNDQOVKIRGFI	EPGEVEARLLEHPAHEAVLVVDGQDKRLVAVYVAEA
Db	3096	PDARLYRTGDLARYLPDGNLEFLGR	NDQOVKIRGFI	EPGEVEARLLEHPAHEAVLVVDGQDKRLVAVYVAEA	4235	QY	110	-----	109
QY	64	-----	-----	-----	63	Db	4236	DEMLVNLRLDLSAVLPDYVMPAA	FVCLDAFPLTPNGKLDRRALPVGEBEDFARQIYAVP
Db	3156	GDCQDKELVAYVAEANEELVNLRL	THLSAILPDYVMPAA	FVCLDAFPLTPNGKLDRRAL	4295	QY	110	-----	109
QY	64	-----	-----	-----	63	Db	4296	SGEMETTLAAIWRELLGVERIS	RYDNFFALGCHSLLAVRMNRI
Db	3216	PAPGAFAFARQVYAAPQGETETLMA	AIWSELLGIEQVSRHDSF	FALGCHSLLAVRMNRI	4355	QY	110	-----	109
QY	64	-----	-----	-----	63	Db	4356	TLMAFAVMSARFDESGEALPAI	IPLSREGQLPLSPAQQRLWLLAQFEGVSEYTHIPLAL
Db	3276	RNAGLTLAVRDLFQSPVLSFEAQ	TGQSPVVVVV	PANVITPALTTLTPEMLPLIDL	TQSEI	3335			

QY	110	-----	109	Db	5496	QMAFADRSLLSYLVAVQQVNNRHTLRTAFIWOGLSAPVQVVMRQAPLSVVVELILD PAD	5555
Db	4416	RLRGLDIATWOQALDTLFAHREALRSVFVAVDGPQVBLPAALGWPMMKYDLRNP DV	4475	QY	124	--VYSQ-	127
QY	110	-----	109	Db	5556	GPVYEQLAQRFNPROHRLDLGOAPLLRFVIAQEMDGRWIALQLOHHLIGDHTTLEVMNRE	5615
Db	4476	DAQLASLSTQERAGTPPDLARGPLIRGALVOLANDEHVFLTLHHIIFDGWSVSLMCELN	4535	QY	128	-----	127
QY	110	-----	109	Db	5616	VQAYLTGQBSLPVPAPFRNLVAQVRLGVNOAEHTFFTFDMLAEVDQPTLPFLGTEVHRD	5675
Db	4536	ALYTAFLTGQDPPLSTIQPYDYAAWQWLSABRIQFQSDYWRARLVDA PVLLDLPTD	4595	QY	128	-----	127
QY	110	-----	109	Db	5676	GSQMTQSYRMLPAIILNNRLRSQARLGVSLAALCHLAWAQVLSRTSGQKVPVGTVLFOR	5735
Db	4596	RPRPRQSPTGNILPVSldaELTQSLKRLSEQGVTLFMTLLAAWATVLSRSGQBDLVI	4655	QY	128	-----	127
QY	110	-----	109	Db	5736	MQTGGTDSMGLFINTLPLRLDMDDTPVRGRVQAAHARLAELLEHEHASLAWAQRCSGI	5795
Db	4656	GTPSAGRSQEVESLIGFFVNTLALRMDLSGAPNVUTELLAHVRQTALAAQEHQDLPFEQV	4715	QY	128	-----YNHAKTIT-	136
QY	110	-----	112	Db	5796	QGEVPLFSALLNYRHSALSATSNEIMNGVEFLGAOERTNYPTLSVDEFGNALGLTTQIV	5855
Db	4716	VEIVQPLRLAHTPLFOVIFAWQSNENAEWQLPGLTVLPADQLIDIAKFDLLELSLSEVDG	4775	QY	137	-----	136
QY	113	-----YSTQL-	117	Db	5856	QPFDPERVCYMQQALESIAETLELAPETFVRTLEILPEARTLLKTNWETEIAYSDPL	5915
Db	4776	RIVGYLNYATALFDQATIERQVGYLHTVLRAMVNPQPVGKIDILTVAERKLLLETWSM	4835	QY	137	-----	136
QY	118	-----	117	Db	5916	CIHQLEQOQVEKNPDATALVYEEQYSYAELNTRANLAHQIALGIVPDQVAVICVTRS	5975
Db	4836	TETSPDHGCIHRLFQWAEKNPDATALMYEGQAISYAEINTSANRLAHQLITLGVVPDQ	4895	QY	137	-----	136
QY	118	-----RKP-	120	Db	5976	PTMIVGLLAVLKAGAYVPLAPAYPGERLAHILTDAAPAILLADNVGCAALGEEALSGLT	6035
Db	4896	RVAICVARSPAMIVGLLAVLKAGAYVPLDPAYSGERLGHILVDKAPAILLADNTGRAAL	4955	QY	137	-----	136
QY	121	-----	120	Db	6036	VLDPNLTPKPDSPNPQVLTALTAQHLAYVIYTSGSTGTPKGVMVHHRGLNLVRDKIAQPD	6095
Db	4956	GEKALADLIMLPNILFDQADSNPLIPELTERHAYVWYTSGSTGTPKGVMVHHRGVNVL	5015	QY	137	-----	136
QY	121	-----	120	Db	6096	IHSDSRILQFASLSPDASVWEIMMALGSGACLVIAVDIVRQDPLRLWHVLEQQVWTHACL	6155
Db	5016	ALAQITRFSVDETSRILQFASGFDASVSEIMTALSGGACLVIPTDIIROQPRRLMSYLE	5075	QY	137	-----	136
QY	121	-----	120	Db	6156	TPALLRDGDLPALATITPTVILGGEAPSAAALFQTLCCRRAALFNAYGPTBITVCAATWRCP	6215
Db	5076	KQAVTHAFLPPALFREESDLPVITIKPTLIFAGEAPGSTLFRALCDRVNLFNDYGPTEIT	5135	QY	137	-----	136
QY	121	-----	120	Db	6216	PDYTDLVPIGHPTANTQIYLLNSDGPVPLGAVGELYVGGAGVARGYLNRRBELTAERFL	6275
Db	5136	VCATSWSCPSDYTDWVPIGRPTTNRVYLLDYGQVPLGTGVGELYIGGVGVARGYLNLY	5195	QY	137	-----FR-	138
QY	121	-----AKY-	123	Db	6276	ADPFSEMPGARWYRTGDLARYLPDGNLVFVGRNDQQIKIRGFRIEPEGEIETRLTEYPAVR	6335
Db	5196	PELTAERFLTDPFSNEPDARLYRTGDLARYLPDGNLI FVGRNDQQVKIRGFRIEPOEIEA	5255	QY	139	-----	138
QY	124	-----	123	Db	6336	EAVLARGEHDKRLVAVWVAEENDELANSRLTHLSTILPDYMPAAAFVRLDALPLTPNG	6395
Db	5256	RLMEHPAVSGALVALSDGQDKELVAVVAEPDGLTASLRGYLSAILPDYVMPAAFVRL	5315	QY	139	-----	138
QY	124	-----	123	Db	6396	KLDRLALPADDEAFSRQIYEAPQGETETALAAIWRELLGIELVGRYDSFPALGGHSLLA	6455
Db	5316	NTFPPLTPNGKLORRALPAGADAFARQVYAAPOGETEILLATTIWSELLGIEQISRHDSFP	5375	QY	139	-----	138
QY	124	-----	123	Db	6456	VRMINMAALGIELPLTTLFKSPTLADFAEVMRAQIQEQNNSVSAMLPIISREGALPLSPA	6515
Db	5376	ALGGHSLLAVRMIERLURNAGLTLAVHDLFQSPVLSAPAEVQSGQAVVVPENVITPATTA	5435	QY	139	-----	138
QY	124	-----	123	Db	6516	QORLMLLTQFEGVSQTYHVPMALRLHGQDIAAQQOALNRLFARHEALRSIFISEDGPQ	6575
Db	5436	LTPPEMLPLIELTQTEIDYIVQQMPGGIVNIQDIYALSPLODGLTFHLLTNEGDPVLLVS	5495	QY	139	-----	138
QY	124	-----	123				

Db 6576 VELLPAELGLPMSKHDRLKVFDDADKRLERLCVQEAEPFPLAGGLIQRADDDYV 5635
QY 139 -----
Db 6636 FLITQHIIYDGSAGVLMHLSALYTAFSMGLPDPPLPLAIQYPDYAAQRCWLSAERL 6695
QY 139 -----
Db 6696 QIQSDYWRMSLADAPVLLDPTDRSRPQSQSFAGALLPVNDABLTQSLKRLSEQQVTL 6755
QY 139 -----
Db 6756 FMTLLAANATVLSRSGQEDLVIGTSPSAGRSROEVESLIGFFVNTLALRMDLSDSPNAE 6815
QY 139 -----
Db 6816 LLARVQTALAAQEHQDLPEQVVEIVQPPRHLAHTPLFQVMPFAMQNNEGMEWMPGLTV 6875
QY 141 -----
Db 6876 SPAEQRIDIAKFDLELSLSEVDGRIVGYNLYATAPDQLTIERQVGYLHTVLREMANPQ 6935
QY 141 -PCGRV-----PSMTCL-----
Db 6936 QPVGKIDILTVAERKLLLETWNAOTRYPDQTCIHLRFQOBEKTPDAVALVYEHILSY 6995
QY 152 -----
Db 6996 TELNABANRLARQLIKWGISDERVAVILLERSIELIVQAAILKAGAVVPIPRVDPDR 7055
QY 152 -----SEMLNVS-----
Db 7056 KNWLINDCSAKLLSGIPVDVAIPRPSLTDEMTIEEDHRNLDLPRSSTDLAYIMYTS 7115
QY 160 -----
Db 7116 STGTPKGMVPHRAVRLVINNGYAEIGDPDRVFEANPAFPASTPEVWAPLLNGTLLV 7175
QY 160 -----
Db 7176 IDHTLLTPKEFVQALQYRITVLMISVGLFNRLATALFPVLPOIKTLIVGQDVLDPHYM 7235
QY 160 -----
Db 7236 AQVLRDSPQQLNGYGPSEGTFTTTRITALSPEVSRIPIGRPIANTRVYLLDTYQOP 7295
QY 160 -----
Db 7296 VPQGVTEIYIGDGVACGYLNRPETLAEFLPDPFSDKPDARLYRTGDLARYLPDGNLE 7355
QY 160 ---RND-----
Db 7356 FLGRNDQOVKIRGFRIELGEIARLELLEFPVSEAVVQVMGQDQKRLIAYVTEADEELV 7415
QY 163 -----TGEQG-----
Db 7416 NRLHLLSTILPDYMYPTAFVRLDTPLTPNGKLDRRALPAPGEAFARQVYEAPOGETE 7475
QY 168 -----
Db 7476 ILLAAIWSELLGTEQISRHDSFPALGSHSLLAVRMIERLNVGLTLAVRDLFQSPMLSAP 7535
QY 168 -----
Db 7536 AQTVGOSPPVVVPANVITPATTALTPEMLPLIELTQLEIDHIVGQVPGGMANIQDIYALS 7595
QY 168 -----
Db 7596 PLQDGLTFHLLANEGDPYLLVSQMTFDDRALDRYLAQVQVNVNRHDLRTAFVWQGLS 7655
QY 168 -----
Db 7656 TPAQVVMRQAPLSVTKLTLPDVPDGPVRDQALQFDPRRYRIDLQRAPLLHFVVAQEDGR 7715

QY 168 -----CQNFTTFNDM-----F 178
Db 7716 WSVLQLLHHLIGDHTTMEVMHGEVOAYLVQEBENLPVPFPRNLVAQVRLGVSAQBHTRF 7775
QY 179 F-----
Db 7776 FTDMLAEVDEPTLPFGLTEVHRDGSQVTSHERMLPVRLNDRLNQARRGLVSLATLCHLA 7835
QY 180 -----
Db 7836 WAQVLSRTSGQDKVVGTVLFGRMQAGQGDSSMGLFINTLPLRLDMDDTTPQESVRAAH 7895
QY 180 -----
Db 7896 SRLAGLLEHEHASLALAQRCGIPGEAPLFSALLNRYHSAQPVTTDATISGIEFLGQER 7955
QY 180 -NVP-----
Db 7956 TNYPFVLSVEDVGHALGLTAQIAQIDPERICDYMQALESALVALEQHEPEIPVRTLNL 8015
QY 183 -----RWT-----
Db 8016 PTAEXKLLLETWNAETSYPDHCCIHFQEWAEKSPDATALVYERQSFSAELNNRANR 8075
QY 187 -----
Db 8076 LAHQIALGVIPDQRVAVCVSRSLAMVGVAVLAVLKAGAVVPLDPVYTGERLTHILTDA 8135
QY 187 -----
Db 8136 PAILLADNVGRDVLGEDALAGLTVLDPNSPPQPOSNPQVPAALTAQHLAYVYTSSTGT 8195
QY 187 -----
Db 8196 PKGVVVEHRIILFEATEFTWTFNRQDIWCLFHSIAPDFSVMELGALRYGAKLVLP 8255
QY 187 -----
Db 8256 AIARSPQELHQFCQHGITVLNQTPSAFKAFIASYIANPLDCLRYIIFGGEALDPSILK 8315
QY 187 -----
Db 8316 PWYALREETLPQLVNMYGITETTVHYTRALARHDEQITSPITRIPDLTLYLLDKYSQ 8375
QY 187 -----KLYVG-----
Db 8376 PVPGLAVGELYIGGAGVARGYNRPETLAEFLPDPFSSEPDARMYRTGDLARYLPDGNL 8435
QY 192 -----
Db 8436 AFLGRNDQOVKIRGFRIEPEIEARLTEHPAVHEAVVQVMGQDQKRLVAVVVAEVDKEL 8495
QY 192 -----
Db 8496 INLRTHLGAILPDYMPAAFCMDAPLTPNGKLDRRALPVGENAFARQVYAAPQGET 8555
QY 192 -----
Db 8556 ETLLATIWSELLGIEQVSRHDNFPFALGSHSLLAVRMIERLNVGLTLAVRDLFQSPVLSA 8615
QY 192 -----PTKV-----
Db 8616 FAQTAGQSLEVVVPTNVITPVTALTPEMLPLADLTQSEIDHIIAQVSGGVANIQDIYGL 8675
QY 196 -----
Db 8676 SPLQDGLTFHLLANKGDPYLLAGQVNFANRDLNLYLAQVQVNVNRHDLRTAFVWQGL 8735
QY 196 -----
Db 8736 SAPVQVMRQAPLSVTELTLPDVLGVCQLAQHFPDRRYRIDLQRAPLLHFVAQEDDG 8795

QY 196 ----- 195
Db 8796 RWSALQHQHLLIGDHTTMEVMOGEIQAYLAGOENLPVPVPPFNLVQAARLGVQAAHTR 8855
QY 196 -----NVDSQTIYF----- 204
Db 8856 FFTDMLSEVDEPTLPFGLMEVHRDGSQVMESHOMLTPELNDRLSQARRMGVSLATLCHL 8915
QY 205 ----- 204
Db 8916 ANAQVLSRTSGQKVVGTVLFGMQAGQGADNGMGLFTNTLPLRLDMDTTPVRDGVRLA 8975
QY 205 ----- 204
Db 8976 HTRLAELLEHEHASLALAKRCSEVPRETPLFSALLNVRHSAQPVTTDVTISGIEFLSGQE 9035
QY 205 -----IGLT----- 208
Db 9036 RTNYPVLVSVEDFGHALGUTVQIVQIDPKRVCGYMQALESAAALBQOQETPVRLNLI 9095
QY 209 -----AL-----LRYAQRN----- 218
Db 9096 LPEAEKXLLLATWNAQAPYPDQSCIHRLPEQQAENTPNAIALVYGHILNVAELNARAN 9155
QY 219 -----C 219
Db 9156 QLAHWLIGQGVKPDNRVAVILLERSVELVVAQLAKAGAVYVVIDPRVDPERRYWLISDC 9215
QY 220 T----- 220
Db 9216 SAKLLTDTPIDIAIPRCLADEMGAIROEDNRNPDLPSSTELAYIMYTSGSTGTPKGV 9275
QY 221 ----- 220
Db 9276 MYPHRAVRLVINNGYAEIGPDRAFEANPAFDASTFEVWSPLNGVLVIDHATILM 9335
QY 221 ----- 220
Db 9336 PKELVQALQTHRITVLWLSVGLFNRLAAELSPAPQKILIVGGDVLDPHVIRQVLRDNP 9395
QY 221 ----- 220
Db 9396 POQLLNGYSPSEGTFTTTVYRITALSPENVVRPIPIGRPIANTRVYLLDTYGHVPVQQAIGE 9455
QY 221 ----- 220
Db 9456 IYIGDGVVGYLNRPELTKERFLPDPFSDPEPNARLYRTGDLARYLPDGNLEFLGRNDQ 9515
QY 221 ----- 220
Db 9516 VKIRGFRIPEGETEAQLKHPAVSEALVLTGNGQDKRLVAVYVVAEADMLVNLRLTHLS 9575
QY 221 ----- 220
Db 9576 MILPDYVPAAFVRLDTPFTPNPKLDRALPVGBEEDFARQIYAAPLGEMETLLATIMC 9635
QY 221 -----HSFVLVNAAGR----- 231
Db 9636 ELLGVERISRYDNFFVLGCHSLLAVRMNRIRIAGLIEWPLSTLTPFPVLTTPABAISAQF 9695
QY 232 ----- 231
Db 9696 GESDRILPAIVPLSREGQLPLSPAQORLWFLTOFGVSETYHIPLALRLHGRLDIAAWQ 9755
QY 232 ----- 231
Db 9756 TLDTLFARHEALRSFVAVDGPQIPELLPAASGLMRKYDLRDMPDVDAQLASAQEIE 9815
QY 232 ----- 231
Db 9816 TLFDLARGPLIRSSLIQLADDDHVFLLTLHHIVDFGWSVLMRELSSLYTAFLAGQDP 9875
QY 232 ----- 231

Db 9876 LPPLTIQYDYAAQOROWLSAERIQSQSDYWRARLVDAPVLLDLPTDRPRPLQQSFAGAL 9935
QY 232 ----- 231
Db 9936 LPVSLDAELTQSLKRLSEQQGVTLFWTLAAWATVLSRLSQGBDLVIGTSPSAGRQOEVE 9995
QY 232 ----- 231
Db 9996 SLIGFPVNTLALRMDLSDAPNVTELLARVRQTALAAQEHQDLPFEQVVEIMQLPRQLAHT 10055
QY 232 -----NLFRV----- 236
Db 10056 PLFQVMFAMQNNENTEMWRLPGLEVSPEQVFDVAKFDLELNLFEBEGVIGVLYNYATALF 10115
QY 237 ----- 236
Db 10116 DOATTIERQVCYLHTVLRGVMANPQOPVGKIDILITVAERKLLLETTWNAWTEMPYPAHCCIHR 10175
QY 237 ----- 236
Db 10176 LFEQQAESPDATAALMYAGQVLSYAEELNARANRLAHQLIMLGVAPBQRVAICVASSPARI 10235
QY 237 ----- 236
Db 10236 VGLLAVLKAGAYVPLDPAYPGERLVHILTDVAPVILADNTGRTALGKVLAAALTVDLP 10295
QY 237 ----- 236
Db 10296 NSLPQPDSPNQVSALTPRHLAYVIYTSGSTGIPKGVVVEHRLVNLIOEKIVQPEIHFG 10355
QY 237 -----PKYI----- 240
Db 10356 SRMLQFASFGFDAGVWETWMLCSGATLAIADTVRQEPRLVHLYLEQAIHACLTPAL 10415
QY 241 ----- 240
Db 10416 LREGTDLPEMTIRPTLIILGEAPSATLLOALSRRATVFNAYGTEITVCATTWRCPSDYT 10475
QY 241 ----- 240
Db 10476 EGVIAIGRPTVNTQVYLLNTDQOPVPLGAVGELYIGGIVARGYLNRPDLTAERFLADPF 10535
QY 241 ----- 240
Db 10536 SDEPDARLYRTGDLARYLPDGNLEFLGRNDQOVKIRGFRIEPEIEARLIEHLAVSDALV 10595
QY 241 ----- 240
Db 10596 LTLGQDQDKRLVAVYVVAEADDQLVNNLYAHLRAILPDYVMPAAFTRLDTPFTPNKGLDR 10655
QY 241 ----- 240
Db 10656 RALPAPGVEAFARQVVAAPQGEIETFLAAVWRELLIGIEQVSRHDSFFALGGHSLAVRW 10715
QY 241 ----- 240
Db 10716 ERLRAGLTLAVRDLFQSPVLSEFAQTVGQSPVAVPTNVITSATTKLTPEMLPLIDLQ 10775
QY 241 ----- 240
Db 10776 TEIDHIEQVPGGIANIQDIYALSPLODQGLFHHLLASEGDPYLLVGQWVFADRALLDY 10835
QY 241 ----- 240
Db 10836 LVAVQVINRHDILRTAFINQGLSTPVQVWCQALLSITELTLNPIDGPFVRDQLAQFDP 10895
QY 241 ----- 240
Db 10896 RRYRIDLKEAPLLHFVVAQEDGRWSVLQLLHLLIGDHTTLEVMNGEVOAYLAGQEBNLS 10955
QY 241 -----NGT----- 243
||:

Db 10956 APVPRNLVAQARLGSGQAEHTRFTDMLAEVDEPTLPFGLTEVHRNGSQAEHLRMLTP 11015
QY 244 KLNTRKLRK----- 255
Db 11016 ELNRLRSQARRLGVSLLAALCHLAWAQLVLSRTSQAKVFGTVLFGRMQVGEVDSGMGL 11075
QY 256 ----- 255
Db 11076 FINTPLRLDIDETSQVDSVRLAHLRLAGLLEHSHASLALAQRCQVQNEAPLPSALLNY 11135
QY 256 ----- 255
Db 11136 RHNVPAIATDAIISGIEFLSGQERTNYPLVLSVEDEGDLGLTVQIVQIDPBERICGYMQ 11195
QY 256 -----QAPVKQ-----FEKKAKKT 270
Db 11196 QALESALAALEQOPEAPVRLNLPASEKKLLLESWNATQAPYPDQCCHRLPEQAENT 11255
QY 271 ----- 270
Db 11256 PDAIALIYGEHIFNYAELNARANRLARLRIQIOPIERIAVLLERSIELVWAQAILKA 11315
QY 271 ----- 270
Db 11316 GAVYVDPSPDERKHLIRDCSAKLLDIPIDLAIPRFCLADEWGAIRGEDYLNLDL 11375
QY 271 --QSTTTPFSYTSAA----- 285
Db 11376 SRSSTELAYIMVTSGSTGMPKGVMPHRAVRLVINNGVAEIGPDDRVAPEANVPDAST 11435
QY 286 -----I 286
Db 11436 FEVWAPLNGALVVIDHATVLPKPSAQVLRNYRITVLSVGLFNRLAAELSPVFPQL 11495
QY 287 NV-----TTNVTYSITTA----- 299
Db 11496 NILIVGDVLDPHVMAQVLRDPSPPQLLNGYGPSEGTITTYIRIATLPSGVIRIPIGRP 11555
QY 300 -----ARRV 303
Db 11556 IANTRYVLLDANGQVPLGVEGEIYIGDGVACGYLNRPELTAERFLPDPPFSDKPDARLY 11615
QY 304 STSTIA-YRPDS----- 315
Db 11616 RTGDLARYLPDGNLEFLGRNDQVQKIRGFRIEPEGEIARLMBHPAVHEAVLWMSGDQK 11675
QY 316 ----- 315
Db 11676 RLVAIVVAEADMLVNHLRDHLSAVLPDYWPAAFCVCLDAFPLTPNGKLDRRALPAPGEE 11735
QY 316 -FMKSI----- 320
Db 11736 DPARQIYAAPSGEMETTLAAINRELLSVERISRYDNFFALGGHSLAVRMNRIALGIE 11795
QY 321 ----- 320
Db 11796 LPLSTLTFPTLMAFAVMSARDESGEALPAIIPLSREGQLPLSPAQORLWFLAQFEGV 11855
QY 321 -----MATOLR---DLATW----- 331
Db 11856 SETYHIPALRURGRLDIAAWQRALDPLFARHEALRSVFVADGQPVQVELLPTEWGLPMR 11915
QY 332 ----- 331
Db 11916 KYDLRTPDQDAQASFSVQEASVDFLARGPLIRSLTVQLADNDHVFLLTOHHIVSDGW 11975
QY 332 ----- 331
Db 11976 SLGVKSELETLYSAYLNQOPDPLLTQIYPDYAAWQCWLSSAQIQSQSDYWCCTKLAD 12035
QY 332 -----VYTTI----- 336
Db 12036 VPVLLDPTDRPRPQOSFAGNTLLISLDAELTQSLKLLSEOGVTLFMTLLAAWATVLS 12095

QY 337 -----RYRQN----- 341
Db 12096 RLSQOEDLVIGTPSAGRSRQVESLIGFVFNTIALRMDLSGEPVAELLARVRQTALAAQ 12155
QY 342 ---PF-----CBPSRN----- 349
Db 12156 EHQDLPFEQVVEIVQPSRRLAHTPLFQVMPANQNNTENWRPLGLVVSPVEQAFDVVKPD 12215
QY 350 -----RTAVSEFMKN----- 359
Db 12216 LELNLFVEGRITGYLNYATALFDQVTIERQVGYLHMLREMVANPQQTVGKIDILTVAE 12275
QY 360 -----THVL----- 363
Db 12336 LIALGVIPDQRVAVCVSRSPAMVGVLAVLKAGGAYVPLDPVYTGERTLHILTDAAPAIL 12395
QY 364 ----- 363
Db 12396 LADNVRDVLGEDALAGLTVLDPNSLPDQPDSPQVPGLTQAHLAVIYTSGSTGTPKGV 12455
QY 364 ----- 363
Db 12456 MVEHRHLRLFDATESWYRFNRQDIWCLFHSIAFDPSVMELMGALRYGAKLVLPHPAIAR 12515
QY 364 -----IRNETP-----YTIYG----- 374
Db 12516 SPQELHQFVCQHGVTVLNQTPSAFKAFIASYVANPLPDCRLYIIFGGEALPSMLKPWYA 12575
QY 375 ----- 374
Db 12576 LREETSPQLVNNYGITETTVHTVYHALARHVEQTTSPIGTRLPDLTLVLLQYQGPVS 12635
QY 375 ----- 374
Db 12636 GAVGELYGGAGVARGYLNRPBELTAERFLTPDPSHIPDARMYRTGDLARYLPDGNLEFLG 12695
QY 375 ----- 374
Db 12696 RNDQVKIRGFRIEPEGEIARLMEYPAVREAVVLADDDQDKRLVAVVVAEVEELINRL 12755
QY 375 ----- 374
Db 12756 RTHLGAILLPDYWPAAFVRLTFPLTPNGKLDRRALPAGEDAFAPQVVAAPQGETETLL 12815
QY 375 -----TLDMSLY----- 382
Db 12816 ATVWSELLGIEQISRDHGFALGGHSLAVRMIERLNRVGLTLAVRDLFQSPVLSAFAQT 12875
QY 383 ----- 382
Db 12876 VQSPAVVVPANVITPVTTLTPDMLPLDTPQSEIDHIVEQVPGGMANIODIYALSPQL 12935
QY 383 -----YNETMVE----- 390
Db 12936 DGIPLFHLANKGDPYLLVSQTAFFVDRSLDSYLAVQQVINRHDIILRTAFIWOGLSTPV 12995
QY 391 ----- 390
Db 12996 QVWVQAPLSVTELTDPIDGVPYEQLIQRPDPVRHRLDSQAPLLRFMVVAQETDGRWIA 13055
QY 391 ----- 390
Db 13056 LOLOHLLIGHDTTEVMNRVQAYLTGREENLSAPFPFNLVAQVRLGVQAEHTRFTFD 13115
QY 391 ----- 390
Db 13116 MIAEVEDEPTLPGLTEVHRDGSQVTELHRLMPTLIINRLRSQARRLGVSLLAALCHLAWAQ 13175

QY	391	-----NKTASDNKTTPT--	14256	FFVNTLALRMDLSGAPNVTELLARVRQTALAAQEHQDLFFBQVVEIVQSPRQLVHTPLFQ	14315
Db	13176	VLRSISQEKWVGTVLFGRMGGEGADSGMGLFVNVTPLRLDINNTPVQDSARLTHTRL	411	-----RTF--	413
QY	404	-----		-----	
Db	13236	ABELLAHEHASLALAORCSVSGETPLFSTLLNVRHSAQPI TVDETISGIESFNGQERTNY	14316	VNFQWQSNENRDMPLPGLTISPVGSRFDIAKFDLELNLSMDGRIVGVINYVATTILFDQST	14375
QY	404	-----		-----	
Db	13296	PFTLSVEDFGHALGLTAQVVQPPDKRIGCYMQQALESIVALEQAPETPVRALNITLPEV	14317	-----	413
QY	404	-----		-----	
Db	13356	ERRLLKTNWATETVYPKQLCVHQLFPEQQVEKNPDATALVYBEQTFSYAELNTRANLRAH	14376	VERQVGLHTVLREMTANPOQPVGKIDILTVEERKLLLESWNATQTRYPDQTCIHRLEFQ	14435
QY	404	-----SPSM--	414	-----	413
Db	13416	QLTALGIAPDQORVAICVTRSPTVMVGLLAVLKAGGAYVPLDPAYPGERLAHILTDAAPAI	14436	QTEKTPDAVALVYEBRILSYAELNARANRLAYRLIERGVQSDRRIAVILERSIELVVAQL	14495
QY	408	-----		-----	
Db	13476	LLADNVGCAALGEEALSGLTVDPNLTLPKPDSPNQVTTALTAQHLAYVIYTSGSTGTPKG	414	-----IDP--	416
QY	408	-----		-----	
Db	13536	VNVHRGLINLVMDKIAQDVHSDSRILQFASLSPDASVWEIMVALGSGASLI IAVDIVR	14496	AILKAGAVVYVFDPRMPDERKNWLISDCSAKLLLSLILNVLAVLQFCLDDEMDAIRQEDH	14555
QY	408	-----		-----	
Db	13596	QDPLRLMHYLEQAVTHACLTPAFRLERTDLPMTIRPTVILGGEAPSVMLLQALSRRAT	417	-----	416
QY	408	-----		-----	
Db	13656	VFNAYGTEITVCATTWRCPDPTDMLVFIHPTANTQIYLLDSYQCPVPLGVVGELYVG	14556	CNLDLPRSSTELAYIMYTSGSTGTPKGVMPHRAVVRVRLAINNGYAEIGADORVAFEANPV	14615
QY	408	-----		-----LW--	418
Db	13716	GVGVARGYLNRPETLTAERFLSDPSPDEPNARLVRTGDLARYLPDGNLEFLGRNDQVKIR	14616	FDASTFEVWAPLLNGALVVIDHATLLTPKBFVQALQTYRITVLWLSVGLFNRLATALSP	14675
QY	408	-----		-----	
Db	13776	GFRIEPEGEIARLMEHPAVQEAUVLVLDGDEQNKLRYIYVAEADAELVNLRLTHLSAILP	419	-----	418
QY	411	-----		-----	
Db	13836	DYMPAAAFVRLNAPFLTPNGKLDRRALPVPGEAFVVRQIYAAPSGEMETTLAAIWRELLG	14676	VLPQIKTLIVGGDILDPHVIAQVLBDGPPQQLNGYGPSEGTFTTTTYRITALSPEVSRI	14735
QY	411	-----		-----	
Db	13896	VERISRYDNFFALGCHSLLAVRMMNRIAALEIBELPLSTLFTFPALADFAEAINTRLNEQS	419	-----	418
QY	411	-----		-----	
Db	14016	LFSRHKALRSVFVAVDQPOQVELLPAASGLPMKKYDLRNAPDADAQIASLQAQETEVPEFN	14736	PIGRPIANTRVYLLDTYQOPVPGVTGTETTLAAIWSSELLGVEQVSRHDSFPALGCHSLLAVRMIERL	14795
QY	411	-----		-----	
Db	14076	LTRGPLIRSSLVQLANDEHVFLTLHHIIFDGWSVSLMRELNTLYTAFLAGQDPLPLP	419	-----	418
QY	411	-----		-----	
Db	14136	TIQYPDYAAWQROWLSAERIQSQSDYWRTKLADAPILLDPTDRPPRPOQSFAGNILLVS	14796	DARLYRTGDLARYLPDGNLEFLGRNDQVKIRGFRIEPEGEIEVRLMEHPAVSEAVVLAQG	14855
QY	411	-----		-----DYL--	421
Db	14196	LDAELTQSLKRLSEQQGVTLFWTLLAAWATVLSRLSQDDVVIGTIPSAGRSROEVESLIG	14856	DSQDKRLVAYVAEANEELVNLRLTHLSAILPDYVMVPAAFVLDTPFLTPNGKLDERRALP	14915
QY	411	-----		-----	
Db	14255	-----		-----	
QY	411	-----		-----	

Db 15336 LPRLDMDTPVQESVQAATHRLAGLLEHEHASLALAQRCGVQNGIPLFSALLNVRHNA 15395
QY 427 ----LDEIRN----- 432
Db 15396 PVIIVTDEVTNGIFQGEQERTNYPFVLSVEDPSNALGLTTQIVQIPIDPERICDYMQAAL 15455
QY 433 ---- 432
Db 15456 SLABALEQKSETPIRLNLILPEABKLLLESNNATQALYDQCCMRLFEQQAEBTPDAI 15515
QY 433 ---- 432
Db 15516 ALMYGEYILYAEINARANRLARRLIGLIGRSDDRVAVLLERSVELVIAQLAILKAGAY 15575
QY 433 --FSLRSP-----YVN----- 442
Db 15576 VPLDLRIPDLRHLWLSDCSAKLLTIDIPVDLIDPLCYFADEVDDIEEBGYLNLDLPRSS 15635
QY 443 ---- 442
Db 15636 TDLAYIMYTSGTGTGKGMVPHRAVRLVINNGYAEIGPDDRVAFEPANPAFDASTFEVM 15695
QY 443 ----LTPPE-----HRAVNLST----- 456
Db 15696 APLNGGTLVVIDHTTLPQEFVQTLQAYRVTVMMLSVGLFNRLAELSTVLPLQKILI 15755
QY 457 ---- 456
Db 15756 VGGVDLPDPAIAQVLRDGPQQLNGYGPSEGTTFTTYRITALSPGTMQIPIGRPIANT 15815
QY 457 ---- 456
Db 15816 RYVLLDIYQVPVQGTGTGIYIGDGVACGYLNRPELTAEFLVDPFSLDPARMYRTGD 15875
QY 457 ---- 456
Db 15876 LARYLPDGNLEFGRNDQKIRGFRIEPEIEARLVEYPTVQEAIVLVLGDGQDKRLVA 15935
QY 457 ----SNSL----- 460
Db 15936 YVVAQDEGLANSHTYLSAILPDYMWVPSVFRVDDPPLTPNGKLDRLHALPAGNEAPAR 15995
QY 461 ---- 460
Db 15996 QVYEAPOGEMETVLVAIWCELLAEIQVRYDNFFALGSHSLAMRMINLAANHGLICTLN 16055
QY 461 ---- 460
Db 16056 DLFQFPVLSELAAMTSDKLSQPRNSAISVRSBGTGLPLFFVPSGMGDYSYVFLGSQLH 16115
QY 461 ----W----- 461
Db 16116 LGYPIVALPWSISEBPMSTMERQATMFMKAVQEPGYRICGYSSGGILAYATAQQL 16175
QY 462 ---- 461
Db 16176 LYSSEVNVFLGLIDTAPHYRQRITPKLQFFIELARQAEDGYIEEMAALYRRIDELNL 16235
QY 462 ----W----- 462
Db 16236 VQFIETAQKALYANLADLIAKRWEQIANYAQIVRDYEPQVSAITLHQFYAMESSPSV 16295
QY 463 ----WLO 465
Db 16296 PVVIDTKIMDIEPQPLNMGSSLGWAQ 16321

RESULT 7
ID ABB84277
XX ABB84277 standard; protein; 11300 AA.
AC
XX ABB84277;
XX

DT 08-NOV-2002 (first entry)
XX Adenovirus Adilp associated protein SEQ ID 2.
DE
XX
KW Genome; cancer therapy; vascular disease; cytostatic; vulnery; cancer;
KW liver; lung; breast; prostate; bladder; glioblastoma; neuroblastoma;
KW medulloblastoma; hepatoma; adenoviral vector; gene therapy.
XX
OS Mastadenovirus.
XX
PN WO200253759-A1.
XX
PD 11-JUL-2002.
XX
PF 04-JAN-2002; 2002WO-SE0000013.
XX
PR 04-JAN-2001; 2001SE-00000035.
PR 08-JAN-2001; 2001US-0260358P.
XX
XX (WADE/) WADELL G.
PA (MEY/) MEI Y.
PA (SEGE/) SEGEMAN A.
PA (SKOG/) SKOG J.
XX (LIND/) LINDMAN K.
PI Wadell G, Mei Y, Segerman A, Skog J, Lindman K;
DR WPI; 2002-590637/63.
DR N-PSDB; ABQ76121.
XX
PT Novel DNA sequence encoding adenoviral proteins, preferably fiber protein
PT and the hexon of serotype 11 adenovirus for gene therapy of vascular
XX diseases and cancer therapy.
PS Claim 1; Page 52-97; 102pp; English.
XX
CC This invention describes a novel purified and isolated deoxyribonucleic
CC acid sequence representing the Adenovirus type 11p genome. This genome
CC can be used to construct a vector for gene therapy, cancer therapy or
CC therapy of vascular diseases. The product of the invention also has
CC cytostatic and vulnery activity and can be useful for gene therapy of a
CC patient suffering from cancer such as liver, lung, breast, prostate and
CC bladder cancer, for the transfection of human cells such as cells of
CC neural origin, glioblastoma, neuroblastoma, medulloblastoma, hepatoma,
CC breast cancer, endothelial cells and dendritic cells, and for infecting
CC human endothelial cells or hematopoietic progenitor cells. The invention
CC is also useful in vascular surgery, for the formation of neo-intima in
CC vascular prosthesis, such as vascular stents and in the prevention of
CC scar formation in surgical applications. The adenovirus vector described
CC in the invention exhibits a high tropism towards endothelial cells and
CC prevents the formation of neo-intima and microangiogenesis. This sequence
CC represents a preliminary amino acid sequence associated with the
CC Adenoviral type 11p Adilp genome described in the disclosure of the
CC invention
XX
SQ Sequence 11300 AA;

Query Match: 23.7%; Score 580; DB 5; Length 11300;
Best Local Similarity 2.2%; Pred. No. 0.00014;
Matches 233; Conservative 94; Mismatches 137; Indels 10068; Gaps 73;
QY 2 GRKWN----- 7
Db 84 GCKKLLIPARKLNEEVFFIMWYLVQGVFVQGVDFDPLRGGFYRVFVFNFRVPCQSLL 143
QY 8 ----- 7
Db 144 FLRRCOLIARVFIPOGLCOEATLEQREBFSPLRKQFNKNKMRDLRFLPQEIISAETGNE 203
QY 8 ----- 7
Db 204 ILELVVHALMGDDPPFPVQLFEPPTLQELYDLEVEGSEDSNEEAVNGVFTDSMLAANEG 263

QY 8 -----VRDVPKM----- 14
Db 264 LE LDPPLDTPDTRGVIVESGTVRKLPDLSSVCDLHCYEDGDFPPSDEBDEHEKQSQTA 323
QY 15 -----FV----- 16
Db 324 AGEVKAANVGFOLDCELPFGHCKCEFRKNTGVKELLCSLCYMRTHCFYISKCVK 383
QY 17 -----LISISFLL----- 24
Db 384 IRNMLFETCILSVFVLLIIGPVSDADESPSTTSPPEIQAPVPDVVRKPIPVKCLKPG 443
QY 25 ----- 24
Db 444 KRPAVEKLEDLQGGDPLDSTRKPRQVFIHVRVLRQYLCDSSAMKYVNCSLVFIWF 503
QY 25 ----- 24
Db 504 AGTVYKQTCVVSSSELAFTHGGLGHGRPEDATVRFGRSLRFLLEILVRISGSPDKTG 563
QY 25 ----- 24
Db 564 LTRIKVUGRLPRTFSSFGSGSLRKSFI SPRLFNPRNCCCPSFYIRMDPADSFOQGI 623
QY 25 ----- 24
Db 624 RFGHSHSIVENMEGQDEDNRLASAAFGCSGNPEASTGHASGGGTARGQPESRPG 683
QY 25 ----- 24
Db 684 PSSGGGVADLSPLOVLVTGSTGRDGRVKRERASSGTDARSELALSMSRRPETIW 743
QY 25 ----- 24
Db 744 WHEVQKEGRDVSVLQEKYSLEQVKTCLWLEPEDDWEVAIKYAKIALRPDKQYKITERIN 803
QY 25 -----VSPINCKV----- 32
Db 804 IRNACYISGNABEWIDTQDAVIRCCMDMWPVGVMEAVTFVNVKFRGDGNGIVFMA 863
QY 33 -----MSKALYNR----- 40
Db 864 NTKLILHGCSFFGNNTCVDAMQVQVGRCSFYACWIATAGRTKSLSLKCKCIFORCNLG 923
QY 41 ----- 40
Db 924 ILNEGEARVHCASTDTGCFILIKGNASVKHNMICGASDERPYQMLTCAGHCNMLATVH 983
QY 41 ----- 40
Db 984 IVSHQRKKWPVFDHNVMTKCTMHAGRRGMFMPYQCNMNHVKVLLBPDAFMRSLTGIFD 1043
QY 41 ----- 40
Db 1044 MNQIWKILRYDDTRSRVRACBGGKHARFPQVCDVTDLEDLPHLVIARTGAEGSGSE 1103
QY 41 ----- 40
Db 1104 ETDGEYENFGVCSGQIEKFPVSVLQSVETILLRGESSALIQVSHPGQEFVRMLWD 1163
QY 41 ----- 40
Db 1164 LLWMEDPSNPPIQRPMLLVLHLWTQLQPPPLLPLTLCLEWVTMEASWLIPLPLITLL 1223
QY 41 ----- 40
Db 1224 PLRTSVLSPWSPWRLPNVWVNFSLRWSCEYKLSLSARQSLNKKIPESMNKTSLLINQ 1283
QY 41 ----- 40
Db 1284 VFLFHSRTVCRPPPIISIENSVDFFQDPIEVGLNVHGHVFGVIEIPLKGFMLRGSVV 1343
QY 41 ----- 40

Db 1344 NHPVITRSCQVVLHNI FKADCHRALGVGVYKPVELGWVHSGVNYHFLDVGNGIAKIPS 1403
QY 41 ----- 40
Db 1404 WYHVMKHODGVSGTFRKPIVOLGWKSVKEKFGDTLVSSKIFHALIHDSNGAVSGAGKH 1463
QY 41 -PW----- 42
Db 1464 VPWVHIIVMFLSIIISHFNEFGAESTRLGYECSFGPRSIIVPLTDLHFFPQFRGWNHVL 1523
QY 43 ----- 42
Db 1524 GGYEKHRFWGGDLQISEQLEATSGGAINDSYGLQVVVGTTATAVFSKQGGHLVHHFPY 1583
QY 43 ----- 42
Db 1584 MHI FPHQIHEALSSKPLGKVFORFOTVSHGHFGESELLQKFSVPQFSDVPYGISIQOTSSF 1643
QY 43 ----- 42
Db 1644 RQFGRLLBGRWRWASAAARVRSFQGLSVRVVVTVKGCAPAWALARVPRLLILLVNF 1703
QY 43 ----- 42
Db 1704 CRLAPCMSAKQFTMGSSLSASAAWPLARSLPLEVFLHTQYRHFSAYNLGARKTDSGEVAS 1763
QY 43 ----- 42
Db 1764 APQEAQTVSHSTSVQVKSGLSGKTSFPYPFLMRFLPLVMSMSCPRVTNRLSVSPTDFTGL 1823
QY 43 -----RGLVLSKI----- 50
Db 1824 PSSGVPRSSYRNSDHTKARVQASTKEAMWEGRLSTRGSTFFSKVCKHMSPSSTRNV 1883
QY 51 -----GKYKLDQ----- 57
Db 1884 IGLVFTPGVAGVGKAVLCSSSLSGSLSRNVSCWGRYLSKAGMTSALRLSVSKNEE 1943
QY 58 ----- 57
Db 1944 DLILTVPMPMPRFSIWSENTIIFLSSLVANDPYRALDKSLANDRMVWVFFLSARSLA 2003
QY 58 ----- 57
Db 2004 AMLSWTYSRARHFHSGKIWNSSGTLTCHPRLCVKIKSTLVATSPRRGSLVQOQLPPL 2063
QY 58 ----- 57
Db 2064 EOKGGSGSISSSGGSGASMWKIPGSKSLSKLWGVGSSKAICHRAASARSYGLRGLPHGM 2123
QY 58 ----- 57
Db 2124 GWVSABAYMPQMTWGSSKMPVGHRRPPLILARTSYSCDGSPPKLVRLGFSVLTIW 2183
QY 58 ----- 57
Db 2184 RQKVALEEMVGLKMLKWAGRPTESLT KWADSSLVTSSTSAVTSTSRQSSVSMMSPGWFFPS 2243
QY 58 ----- 57
Db 2244 HSSRLRYSRSFYQSSGNFSLSARDPSMNLITALGOQPFSTGREYAAAFSEAVRAKVS 2303
QY 58 -----LKEILRQ----- 65
Db 2304 LTMULRNWYLSMSSQPCSQSWKSTRFLAGLKAKVTSLKRLPALGIKLVMRKGCOT 2363
QY 66 -----LETITISK-----YN----- 75
Db 2364 SARLLITWAARTISSKPLMLCPTMYNMRKGVPLTGLSSKVSRSVSGDKACSAHSCR 2423
QY 76 ----- 75


```
Db 2424 GFACRNDQDSTASAVCNMSRYKCRPIAIFSGVTOKVLGSCCHRHLSILMARSWMLTR 2483
Qy 76 -----VSKOPV-----KN----- 83
Db 2484 RSSPESFMTSMKGTCLPKDPIQVWSTSVKSLSVRGEPIGKWNWISCHQLEDWLLMKKF 2543
Qy 84 -----LTMN----- 87
Db 2544 LRRAEHSCLLYRRPOSQRTGCSMSCTWLPLTRNFGKRPDGCISCSSIFAVSACSS 2603
Qy 88 ----- 87
Db 2604 SVSMVWMLTSPRGQVQTSAREGRSRTARRLELSVLRCGLRVGRDRRLTCMIFSR 2663
Qy 88 -----TEFPQYILAG----- 98
Db 2664 CGRFRWLISTGSFVETSMACRVPCLGATTVPLFLLIGGSLASCMRLSGDGDARRAA 2723
Qy 99 ----- 98
Db 2724 AVVPDPRAMLVVARRRRARSAGSCTALELRAPRVDRLVSDVSGKLPAPATKRVOQNQR 2783
Qy 99 ----- 98
Db 2784 YRQLVSVFLRHQSCFGRSPPTARFLPDLRDLFRRWPRGHRYGPVGRMHSCPPR 2843
Qy 99 ----- 98
Db 2844 SRRGCKPRPRSLRASPPERGAPVRWPHSCIGAEGKSVWQCVRRENTSIVSAAPRH 2903
Qy 99 -----PIQN----- 102
Db 2904 RPELGAPWRPRFRONKTGSFARTRSIPREDGVLMPVLRVLRPGSLPLLSLEPL 2963
Qy 103 ----- 102
Db 2964 TSLRLQAGAEAGDVGARANGRIVQPLRGCGANFQRRGRBRAVAEKHRRASPSGDWE 3023
Qy 103 ----- 102
Db 3024 VLRLGGRGRLYILLIGPLHAEISCQDPRDLKTRFRKRLTSHSHKVGVRLLVGGGYVFG 3083
Qy 103 ----- 102
Db 3084 LGLLLHLLGKVRCCWNSRQFDGWWGAPGLWVRLAGYAGDWPFPKHPYDQIDLCSSL 3143
Qy 103 ----- 102
Db 3144 AAVLRALLPHFPCHAYVQIRALVVPVPSQLRFLRRGWLAVLGGWLESHQNPOSQKLLY 3203
Qy 103 ----- 102
Db 3204 WKGHSWPLTSLSDQAGARCIGANRRGRCRCRCPDTGTLENAAVVGGREAILVLE 3263
Qy 103 ----- 102
Db 3264 ROGRGLPTGDSRRCTWTSRFLARKPETRVRGSKCVAASSLARFDQAGSHCSIDTE 3323
Qy 103 ----- 102
Db 3324 KMKAFSDSTPPGGTTGNVAVYPGSRLVLEPAGAAANVIALPSRPSLQKSRINRVLLV 3383
Qy 103 ----- 102
Db 3384 SEWQGESYFFPPAAQWHPVLRQWRPQQPPSQOQQOQSQKAVPATTATAVSGADSPM 3443
Qy 103 ----- 102
Db 3444 IWTWKAKDWHVVRHPSGHIREFNKKILARRMCPNRTYLEAARRRRCCLPALTRVVS 3503
Qy 103 ----- 102
Db 3504 CVTWTEDECCGTRISKLMKQSVLPFGHTWLOPTLYRLTSRQRKSVTSKSLLIIMCEPLP 3563

Qy 103 ----- 102
Db 3564 AKKLPLVCICGIWKLFRLLANLPSCFWCMTAETMLSERCTSPNPRGDCMILST 3623
Qy 103 ----- 102
Db 3624 FYVSCRSAGAWPRRWLPISITRPAWENITLAKSTRLHTFPTRRRMGSTCARSRSPAMIL 3683
Qy 103 ----- 102
Db 3684 GCIAWTECIARLAPAGASATGCTVCKELLEPRVITSTWELTCSGLVAGLAPRRQ 3743
Qy 103 ----- 102
Db 3744 DVSFLTKRMRKARRKRASTWKTGTTRVFCMEQAPDPAMRAALOSQPSGINSDDWTOA 3803
Qy 103 ----- 102
Db 3804 MORIMALTTRNPEAFRQQPQANRLSAIMEAVVPSRSNPHEKVLATVNALVENKAIKRPDE 3863
Qy 103 ----- 102
Db 3864 AGLVYNALLERVARYNSSNVQTNLDRMITDVBREAVSORERFORDANLGSILVALNAFLSTQ 3923
Qy 103 ----- 102
Db 3924 PANVPRGQDYTNFLSALRLMVSEVPQSEYQSGPDYFPQTSRQGLQTVNLSQAFKNLKG 3983
Qy 103 ----- 102
Db 3984 LMGHAPVGERATVSSLLTPNSRLLLLVAPPTDGSIDRNSYGLVLLNLYREAIGOSQV 4043
Qy 103 -----YSIT----- 106
Db 4044 DEQTYEITQVSRALGOEDTGSLEATLNFLLTNRSQKIPQOALTAEERILRYVQOSVG 4103
Qy 107 ----- 106
Db 4104 LFLMQEGATPTAALDWTARNMEPSMYASNRPFINKLLDYLHRAAAMNSDYFTNAILNPHW 4163
Qy 107 -----YLN----- 109
Db 4164 LPPPGFYTGYDMPDNDGFLDDVDSDVFSPLSDHRTWKKEGDRMHSSASLSGVMGAT 4223
Qy 110 ----- 109
Db 4224 AAEPESASPPSLPFLSHSVRSSEVRISRPISLMGEESEYLNDSLLRPAREKNFPNNGIES 4283
Qy 110 ----- 109
Db 4284 LVDKMSRWKTYAQDHRDEBPGINGITSRASRRRQRHQRGLVWDDSDSDSSVLDLGR 4343
Qy 110 ----- 109
Db 4344 GRGNPFAHLRPLGGLMKIKKLTAMATSVRSFPFIIICVYNEASRARRSGVSGSSS 4403
Qy 110 ----- 109
Db 4404 FVRERDAAMAGGGDAIPTGGSLSASAPGTGGOKQHSFGTGTSTVRYHVQVSGQQVG 4463
Qy 110 ----- 109
Db 4464 GHCPSELSEPQLLDHGGAKQLYPYGSQHPDLHTIATVGRSAKDHAYHAKRERVYVQVS 4523
Qy 110 ----- 109
Db 4524 ACDGVQKTSRRCCSWGILSQAGYFBIYVVRVYFARRQLFSYYDYFDEQCHHRLLESSTWS 4583
Qy 110 ----- 109
Db 4584 AKHWCVRHQELQAGMSRQNDHANSVYVLSLPSHCILTAWLRSGFYRESFEQSPSYOKKTA 4643
```

Qy	110	----	-DPP-	----	111
Db	4644	ISRGDFVFRFRWYSGPLGCRCL	EQERTKSQNRSCYSCCRSGKHSQ	RLYKGRWGRQR	4703
Qy	113	----	-YST-	----	115
Db	4704	QFCANTCSDCR IGRCVNRG	RETHYSTCRKREKLCVGRQNH	LSQLSVFQLWRSRKR	4763
Qy	116	----			115
Db	4764	SAFLDIAHHLCHLRSRAGLL	VASRHDGSCHPPLHTSQLPCG	GCRAVARILKELLORTS	4823
Qy	116	----			115
Db	4824	CVLPAAPVHLAYARLOPLS	BPDFNPSAGAHYHRKSCSHR	SRDPAVAQYPGSPCTD	4883
Qy	116	----			122
			-----OLRPAK-		
Db	4884	RYRQTPHLSLVRQGTGHS	RTARPKPHFLKXKXMSILIS	PNNWTGWLRAFSKMYGGARK	4943
Qy	123	----			122
Db	4944	RSTQHPVRVRGHFRAPW	GALGKRTVRTVDVDOVVAD	ARNYTPAPTSTVDVIDSV	5003
Qy	123	----			122
Db	5004	VADARNYARRKSRRRRI	ARRHRTATTAMRAALLRR	ARRVGRRAWLRAAREASGASGR	5063
Qy	123	----			122
Db	5064	SRRQAAFAAATIADMA	QSRGNVYVRDAATGORVP	TRPPRTKILSSLRCCVPAARM	5123
Qy	123	----			122
Db	5124	SKRKYKBEMLQVIAPE	VYVGPQLKDEKPKRKIRV	KDKKBEEDGDLAEFREFAPRRR	5183
Qy	123	----			122
Db	5184	VQWRGRKVRHVRPGTS	VVFTPGERSSATFKRSY	DEVVGGDDILLEQAADRLGE	FAYGKRS 5243
Qy	123	----			122
Db	5244	RITSKDETVSIPLDH	GNTPSLKPVTLQOVL	PVTPRTGVKREGSDLYPT	WQLMVPKRQKL 5303
Qy	123	----			122
Db	5304	EDVLEKVKVDPDIQ	PEVKVRPIQVAPGLGVQ	TVTDIKIPTESMEVQTEPAK	TATSTEVO 5363
Qy	123	----			122
Db	5364	TDPWMPPIITDAA	GPTRRSRKYGAPALLPN	VVHVHPSIIPTPGVGRTRY	RSRNSTSR 5423
Qy	123	----			122
Db	5424	RRRKTANRRRRRTSK	PTFGALVRQVYRNGSA	EPLTLPRARYHPSIIT	SMLPLPCRY 5483
Qy	123	----			122
Db	5484	GPHLSPSRSHHWP	PRKKLAPKRDVGTRN	ATLQATCYPOALAGFF	TSLNSNYRCCNWRD 5543
Qy	123	----			125
			-----YVY-		
Db	5544	TRHSFRGGSLATTL	TLTKNVIKKKYNGLH	SWSCDVVFLEMEDIN	FSSILAPRHGTPYM 5603
Qy	126	----			129
			-----SQYN-		
Db	5604	GTWSDIGTSQING	AFNAFWSISGLKN	FGSTIKTYGNKAWNS	TGQALRNKLDQNFQOK 5663
Qy	130	----			129
Db	5664	VVDGIASGINGV	VDLANQAVOKKINS	RLDPPPATPGEMQV	EEIIPPEKGGKRRPRDLE 5723
Qy	130	----			147
			-----HTAKTITFRPP	CG-RVPS-	

[illegible]

Db	5804	HGPRARLCHCKYGRTDGGRALVGRLEPTFLLPFFSFWILGSSQDLPVIGSPAPORSCY	5863
Qy	220	-----	219
Db	6864	QGPLYYAGKIYPDRAGPPFCRLRLTFLHVPRLCALAPSHGRKPHEIANWSAKHASFS	6923
Qy	220	-----	219
Db	6924	SPAHPVQSKSTLPSPYPFALFSLPSYTHRGHCVCVPYGCSTHVNANFNKHHFIFLHVS	6983
Qy	220	-----	219
Db	6984	RLCITYLTSRMGSDENQNDPOAIVIRCGTDTWATITRESPTWEPVYRAGCHSTAFWSAAK	7043
Qy	220	-----	219
Db	7044	LQAGQPKSNHNDQCFERESCGTDCSTETPSATDVSRLLPARWDLQCPHPDLQHWQCTG	7103
Qy	220	-----	219
Db	7104	SSCRSAYPMRAPNACGNCNRSAGSVSWPDPVFLDTRLXKHIAKPAGLYPYRYKTSRRT	7163
Qy	220	-----	219
Db	7164	CSKTGLHSRHHSSHSGRHCLFAPHFCPSGFGFWFARDSPRLVLVRSRNPSPRSAPSESY	7223
Qy	220	-----	231
Db	7224	CHAGTSACPHNCHSHEATHTSLYIPYVGRSEKKNVFPABIFPSSCVSCDKLTGCLGA	7283
Qy	232	-----	231
Db	7284	PRLRTGDCACIVRVAQALVRFVYPACTSPSADTSLPCLSPKQTPGASDSQCRQOLL	7343
Qy	232	-----	231
Db	7344	PEGHLRSSQCFFCHPSQRCARAGSNPLQVAPLLFLLRLCLDLAWGVVWSSLASFWGVSE	7403
Qy	232	-----	N 232
Db	7404	EEDCRVPETGRIVTFRSPIDTCRKNLTPHGDRCFSFGAEVIAKAGCGPTWKADWQN	7463
Qy	233	LFRV-----PKYIN-----	GTKL 245
Db	7464	PFRVRCAPCGGLTDFLRGWPLCSQRNRHNGSAIAVNIATSAITSRPPQRGKGAEL	7523
Qy	246	K-----	246
Db	7524	KHSTAOSCHLYPRRGRRISHAEKSERVDRHRARPGLCDTGTRGRVETLSRERKLPK	7583
Qy	247	-----	246
Db	7584	TTSLSPRCWKGEHLPHRARRAPTSSKTVAHSGGCIIGNSAHQCGRAQPLRAPL	7643
Qy	247	-----	246
Db	7644	FTSYSPQTSAKRHLRAKSLKLLSPCCARSTGYLSHLFSKNSLLPRSHPRRCPTQSG	7703
Qy	247	-----	246
Db	7704	TWFTLYSFLGRGSKDLRGSGDGRKCSAKGRKWHGASQSGGIGRRQCOTRSTQAKHRG	7763
Qy	247	NTWR-----	250
Db	7764	HTLIRISRCQAPASHDGHGVPWTHARKSPFRHAPRCLGTSGQANPMAGHRLSQGFGRAS	7823
Qy	251	-----	250
Db	7824	QAYDGRGAGYRRTRVSPFTFLYPRNLAQTRRESALHFTRLCAAGMODIRGTHQPGFLHGY	7883
Qy	251	-----	KLKKAQA-----
Db	7884	SAESPRTKRAAQHEGGSPPLHPRLCLSVFPVNVANRHGCMAMPFRTELEAQALETIS	7943
Qy	258	-----	257
Db	7944	GSVDRVRRARRFRPGRPHLPASQGYFAKRIALYEPHAQFSLFHPGTILYPARHLLRT	8003
Qy	258	-----	257
Db	8004	ALRLCASHLPVPPAPAMESLLPVPSQSLPLGCDRCGERRRLAGVSLPQSVHAPPVP	8063
Qy	258	-----	262
Db	8064	SLQPPVDERNPNRHLIARPPQPRWVFSWAKFKTDPGTVDLRLLAQVCSGRLLPPLNQVL	8123
Qy	263	-----	262
Db	8124	GPITASKRGTRFGLRHHPGGNSGPIASHPKIPPRISTEKGGLPPPPDRRGTOHKVPSCGN	8183
Qy	263	-----	PEKKAK-----
Db	8184	DEKTRSCRRRPQKIWRKIGTVRQRRRTVWRTVWRKTRQRRRKKPPTNSYP	8243
Qy	270	-----	269
Db	8244	RLRQATALPSPLRVVEEPCGVPAVDGTRPDASRTQPALPRVRRIGRDTSPGGGIRMPSS	8303
Qy	270	-----	TOSTTTPYFS-----
Db	8304	PACMSAGATYPSRGATCYSTMGTFRAMFCITVTSTAPTIAKSRQSRQIKTAAATSNRK	8363
Qy	280	-----	VTYSI-----
Db	8364	PAAAVKTYTSAATGGLKITANEPAQTRRLNRIPTLYAIFQOSRQSQBELKIKNRSUR	8423
Qy	297	-----	296
Db	8424	SLTRSLYHKSQDQORTLEDAELFNKYCALTLKEAATALLQKRRLHHPHRSNHAL	8483
Qy	297	-----	296
Db	8484	HVELSTPNCIGGRRLPGLLHPHQAQRAFYDFSSYTRLPKPTFCGTSSYHAPTPSQ	8543
Qy	297	-----	VSTSTIAR-PP-----
Db	8544	KLARRPSVEGKSRSHHCITSSRRPGCRSPNDCRCAVSWRLHPMSQASAYKTDDQRPYP	8603
Qy	314	-----	SSFMKSIATOL-----
Db	8604	AQRRVGLSAMSTTRNLSDCRLREIFLHPSSGCCDFGKVFATPLGRNDRSICGVYS	8663
Qy	330	-----	VYTLR-----
Db	8664	LCLLOPLLRISWALPGRVHTELRRDRVSGRLMSGDAEALSRLRHLDRRCRPRCFAREL	8723
Qy	338	-----	YRNP-FCEPSRNTAVS-----
Db	8724	IBFIYFELPKDHPQAGHVRISIEGKIDSLRQIFSORPVLIERDQNTTTSYICNH	8783
Qy	355	-----	EFMKNTHVLIR-----
Db	8784	PGHESLCLMCTEFKNKIKTLRTAASSTRILQPEERNFSCRPGCLHLHSYQTRSSYT	8843
Qy	368	TP-----	369
Db	8844	TFLQKHFPYFYQNRAPRSSYRKPLGGSGPCARNSCGWACDYSLLPIHTLLHFLSGV	8903
Qy	370	-----	369
Db	8904	VVLVRMGPIVLVLLVLLSLLEPGSANYDPCLDPENCTLTAPDTSRICGVLIKCGWECH	8963
Qy	370	-----	YTI-----
Db	8964	SVEITHNKNTWNTLSTTWEPGVPEWYTVSVRPGDSIRISNNTFIFSEMCDLAMPMSQ	9023

QY 373 ----- 372
Db 9024 YSLWPPSKDNIVTFSIAYCLCACLITALLCVCIHLLVTTTRIKNANKKMPPLSVYRHGF 9083
QY 373 ----- 372
Db 9084 SYISHICQHCHCRSWNSRLYPSRTLLSHRTPNFRHLGQTGKRLLYNLQNKTNNSNL 9143
QY 373 ----- 372
Db 9144 QHTKSYIDCSLQRLLLWLQIQSIKLLGSCYPVENHENAKYKDSIQFSRNFYISHHTR 9203
QY 373 ----- 372
Db 9204 RKKPRFNCNCGSGDGTNNMHAFICLSLQKVSSKTRSPTKAHLISFYTAMVSTTT 9263
QY 373 ----- 372
Db 9264 FLMLTSLATLSARSHLTVTIGSNCTLKGPOGHVFWMRIYDNGWFTKPCDQGRFFCNG 9323
QY 373 ----- YGTLDMSSLYN ----- ETWF ----- VENKTASP ----- 396
Db 9324 RDLTIINVTANDKGFYGTGYDYSLDYNIIVLPSTPAERTTTTFSSSVANNTISPTFA 9383
QY 397 ----- SNKTT ----- 401
Db 9384 ALLKRTWNSSTSHTTSTSTISIIAAVTIGISILVFTITYYACVYKDKHKGDLPLREF 9443
QY 402 ----- PTGFS ----- 406
Db 9444 IFVLFFIYSVMNTNHTKFLHLHHTLCICLRYFSSSHSNPRLYRSICFLCTFCFCYLH 9503
QY 402 ----- 406
Db 9504 LRMHSLPGYFFPYRLDPCANCLPAPSPRIQPKYRGSTHLKPCRLYYQYFCYCFPL 9563
QY 407 ----- 406
Db 9564 SOPQLFVLHONTLENANSNNRGHFLLAIEKKNQPKQILLELISVAPFHYTPYLILAGM 9623
QY 407 ----- 406
Db 9624 LPMHMIHKTORWTFYKTCNIQRIKVNHNPHYSLLLVSTPAEMTETLTTSNSAEDLL 9683
QY 407 ----- 406
Db 9684 DMDGRVSEQLAQLRIRQQERAAKELRDVQIHQCKKGIFCLVKQAKISYEITADHRL 9743
QY 407 ---MGFOR--- 411
Db 9744 SYELGPQOKFTCMVGINPIVITQSGDTKGCIHCSCDSIECTYTLKTLGLRLDLPWN 9803
QY 412 ---TF--- 413
Db 9804 KMINKSLTNQOGLCNFLPAAPHPLPNSGILNPVQRTFSLRGCOILAPLLYPOSSCL 9863
QY 414 ---IDLWDYLD--- 422
Db 9864 SSQMTKRVRLSDSFNPVYFYEDESTSQHPFINPGFISPNGFTQSPNGVLTCLKLPLTTT 9923
QY 423 ----- 422
Db 9924 GGSLLQKVGGLTVDDTNGFLKENISATTPVKTGHSIGLPLGAGLGTNENKLCIKLGQG 9983
QY 423 ----- 422
Db 9984 LTFNSNICIDDNINTLWTVGNPTEANCOIMSSSESDCKLILTLVKTGALVTAFFVYIG 10043
QY 423 ----- SLFLDEIRNF ----- SLRSP ----- 438
Db 10044 VSNFNMLTHTRNINPTAELFPDFTGNTLLSLTKPLNHSKGQNMATGAITNAKGFMP 10103
QY 439 ----- 438

Db 10104 STTAYPFNDNSREKENIYGTCTYATSDRTAFPIDISVMLNRRRAINDETSYCIRITWSWN 10163
QY 439 ----- TYVNLTPPEH ----- 448
Db 10164 TGDAPDEVOTSATTLVTSPTFFYIREDDOIKNFLFIKSIHKIRVWILPPPSHLTEYTNLS 10223
QY 449 -BRAVN-----LSTNSLW----- 462
Db 10224 PRTALNIWIPLDIDMWLDSTFQTVSERANLGSVIDKNPSGFKALSQSNCCGTPESGSR 10283
QY 463 ----- 462
Db 10284 SSGRRTWGIIIRKRYRTIVSHQTHQPLSASLRATAVYGRVHSLKHDNFNSHQLSGAM 10343
QY 463 ----- 462
Db 10344 RAATHSDFTQIFAVGTTHYINVTIISAPAKTHIYNRPCMTIIPKENINMTPPQKHHTH 10403
QY 463 ----- 462
Db 10404 IHDLFWHVHNNLSVPWTTLVNHAQYNLPEPHCOHRSFHALKTLITMTWKNPILSTV 10463
QY 463 ----- 462
Db 10464 NHLRMKNYSGTTTWHASSHNFLLRIKHIPGNRKLQNSKAGTRKTNTVTYTMHSHSIT 10523
QY 463 ----- 462
Db 10524 IWQORVFSHRSSGFILTTWLGSGVRVMSGACRACAOQCHNGVASHSRILYSKTRPWQN 10583
QY 463 ----- WLQ 465
Db 10584 TLFFAFYPAARVPCDSSSTILLSWSKECWLQ 10615
RESULT 8
AAR44929
ID AAR44929 standard; protein; 15281 AA.
XX
AC AAR44929;
XX
DT 16-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 08-JUL-1994 (first entry)
XX
DE T. niveum Cyclosporin synthetase.
XX
KW Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
KW T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase.
XX
OS Tolypocladium inflatum.
XX
PN EP578616-A2.
PD 12-JAN-1994.
XX
PF 05-JUL-1993; 93EP-00810474.
XX
PR 09-JUL-1992; 92AT-00001403.
PR 08-MAR-1993; 93AT-00000437.
PR 29-APR-1993; 93CH-00001310.
PR 04-MAY-1993; 93CH-00001375.
XX
PA (SANO) SANDOZ LTD.
PA (SANO) SANDOZ PATENT GMBH.
PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
XX
PI Leitner E, Schneider E, Schoergendorfer K, Weber G;
XX WPI; 1994-010432/02.
DR N-PSDB; AAQ54386.
XX

PT Isolated DNA sequence - which codes for enzyme having cyclosporin
 XX synthetase like activity.

Claim 1; Page 41-84; 93pp; English.

CC This sequence represents an enzyme which has cyclosporin synthetase- like
 CC activity. This sequence was isolated from Tolypocladium niveum (formerly
 CC known as T. inflatum GAMS). This enzyme catalyses the peptide
 CC biosynthesis of cyclosporins and structurally related molecules. This
 CC sequence may be used for the production of cyclosporin by transforming a
 CC vector containing this sequence in to a recombinant host. This allows
 CC effective production of anti- biotic cyclosporin or its derivatives.
 CC (Updated on 25-MAR-2003 to correct PN field.) (Updated on 16-OCT-2003 to
 CC standardise OS field)

XX SQ Sequence 15281 AA;

Query Match 23.5%; Score 575; DB 2; Length 15281;
 Best Local Similarity 1.7%; Pred.No. 0.00041;
 Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;

QY 1 MG----- 2
 ||
 Db 1 MGAIGQMAYDLANPSRASSISSNRYSEPVQSPAQGLWFLHQLKLGASNDITPAAIR 60
 QY 3 ----- 2
 Db 61 LRGHLDIDALNAASRALTORHETLRTTFKEQGVGVVHASSGLERGLRVDASSRDIAQ 120
 QY 3 ----- 2
 Db 121 LLABEOTMKFLESEPAWRVALLKVAEDHHILSVVHHIISDSRSLDIQOELGELYTAA 180
 QY 3 ----- 2
 Db 181 SQKSISACPLGPIQYRDLTWQNDQVAEQEQLGWIEQLDNNTPAELLTELPRP 240
 QY 3 ----- 8
 Db 241 AIPSGETGKISQIDGSHVHKELLAFCRSQQTAYAVLLAAFRVHPLTGAEDATIGAPV 300
 QY 9 -----RDVP----- 12
 Db 301 ANRDRELENMVAPLATQCMRVVLDDDTFESVLQIMSVMTAEANRDVFFERIVSAL 360
 QY 13 -----KMFVL----- 17
 Db 361 LPGSTOTSRHPLVQLMFALHPAQDTGRARWGFLEAETLQSAAPTRFDMEMHLPFGDDREF 420
 QY 18 ----- 17
 Db 421 ANVLPSTGLFDAEAIHSVVSIFREVLRRGISPAVHVKTMTPLDGLAARDMGLLDIGTT 480
 QY 18 ----- 17
 Db 481 DYPREASVDMFOEQVALNPATAVADASSRLSYSELDHKSQDLAAWLRRLQKPTLIG 540
 QY 18 ----- 17
 Db 541 VLSPPSCETWSPFLGILKAHLAYPLDINVLARIESIISAVDGHKLVLLGSNVPQKVD 600
 QY 18 ----- 17
 Db 601 VPDVELLRISDALNGSQVNLGAKQATAPKPSATDLAYVFTSGTGKPKGVMEHRRGIVR 660
 QY 18 ----- 17
 Db 661 LVKGTNIISPAQAAVPTAHLAIAFDLSTWEIYTPILNGTLVCIEHSVTLDSKALEAVE 720
 QY 18 -----ISISFL----- 23
 Db 721 TKEGIRVAFLAPALIKOCLADRPAAIFAGLDSLYAIGDRFRDRDALHAKSLVKHGVNAYG 780

QY 24 ----- 23
 Db 781 PTENSVVSTIYSVSEASPFVTCVPVGRAISNSGAYMDQDQQLVSPGVMGELVSGDGLA 840
 QY 24 ----- 23
 Db 841 RGYTDSALDKNRFVVVVQIDGESIRGYRTGDRARYSLKGGQIEFFGMRDQOVKIRGRIEP 900
 QY 24 ----- 23
 Db 901 AEVEHALLNSDQVDAAVVIRRQBEPEPAMIAFVTTQGTLPDHLVNINGHVPDGNKSK 960
 QY 24 ----- 23
 Db 961 NDQFAVHVESELRRRLQMLLPSYMPARIVLDHPLNPNKGKDRKALQCSAKTVQKSKL 1020
 QY 24 ----- 23
 Db 1021 VSORVAPRNEIEAVLCEEVRSVLGVEVGITDNFPDLGGHSLTAMKLAARISQRLDIAQSV 1080
 QY 24 ----- 23
 Db 1081 ATVPEQPMADLAATIQRGSTLYSVIPTTEYTPVEQSPAQGLWFLQELNTCASWYVM 1140
 QY 24 ----- 23
 Db 1141 LTVLRGHLDVDALGTALLALEKRHETLRTTPEERDGVGMQVVHSSLMGELRLIDISEKS 1200
 QY 24 ----- 23
 Db 1201 GTAHEALMKEOSTRFDLTREPGWRVALLKLADHHI PSIVMHHIVSDGSLDLRLHELQ 1260
 QY 24 ----- 23
 Db 1261 LYSALRGQDPLSRLEPLIQYRDEFAVMQKDSQKAAHQRLQYWKQLADSTPAELLT 1320
 QY 24 ----- 23
 Db 1321 DFPSPILSGKAGKVPVAIEGSLYDTLQVFSRTHQVTSFVLLAAFRAAHFRLTGSDNAT 1380
 QY 24 ----- 23
 Db 1381 IGVPSANRRPELENNVIGFFVNTQICRITIDENDNFESLVRQVRSRTTTAAQDNQDVPFEQ 1440
 QY 24 ----- 23
 Db 1441 VVSSLPSSSRDASRNPLVQLMFALHGQDLFKIQLEGTEEEVPTTEVTRFDIEPHLYQ 1500
 QY 24 ----- 23
 Db 1501 GASKLSGDIIFAADLFEAETIRGVSVFQEVLRRLGLOQPPTPMTMPLTDGIPELERMGL 1560
 QY 24 ----- 23
 Db 1561 LHMVKTDPYRNMVVDVFOQQVRLSAEATAVIDSSSRMSYASLDQSDQVAAWLRQRLP 1620
 QY 24 ----- 23
 Db 1621 AETFFAVLAPRSCAEVIALFGILKAGHAYPLDINVNPAARLRAILAIEVKEXLVLLGAGE 1680
 QY 24 ----- 23
 Db 1681 PSPEQSPESVIVRIADATSPAGHASLRDGSKPTAGSLAYVIFTSGSTGKPKGVMEHR 1740
 QY 24 ----- 23
 Db 1741 GVLRLVKQTNILSSLPAPQTFRMAHMSNLAFPASIEWEFTALLNGSLVICIDRFTILDQA 1800
 QY 24 ----- 23
 Db 1801 ALEALFLREHINIALPPPALLAQCLTDAATIKSLDLLVGGDRDLTADAALAKALVKSE 1860
 QY 24 ----- 23

QY 98 -----GPI-QNYSITYLWF----- 110
Db 4081 VOQSAPHPDPIVSTKTYGVPQSFQAQRLWFLDQLNFGATWYMLPLAVRLGAMNVHALT 4140
QY 111 ----- 110
Db 4141 AALLALLERRHELLRTTFYBQNGVGMQVNPVVTETLRIIDLNSGDDGYLPTLKKEQTAPP 4200
QY 111 -----DFYSTQL----- 117
Db 4201 HLETEPCMRVALLRLGPGDYILSVWMHIIISDCWSVDVLFQELGQFYSTAVRGHDPLSQT 4260
QY 118 -----RXP----- 120
Db 4261 TPLPIHYRDFALWQKPKTOESEHERQLQYWVEQLVDSAPAEILLDLPFSLISGQAGMS 4320
QY 121 -----AKY----- 123
Db 4321 VTIEGALYKNLEBFCRVHRVTSFVLLAALRAAHYELTGSSEDATIGTPIANRRNPELEQI 4380
QY 124 ----- 123
Db 4381 IGFFVNTQCIRITVNEDETFSLVQQVRSTATAPFAHQDVPEKIVSTILLPGSRDASRNP 4440
QY 124 -----VYSQVN----- 129
Db 4441 LVQLMPAVHSQKMLGELKLENASHSEVVPTEITTRFDLBFHLFQDDKLEGSILYSTDLFE 4500
QY 130 ----- 129
Db 4501 AVSVQSLLSVFOEILRRGLNGDPVPISTILPQDGIVDLQROGLLDVQKTEYPRDSSVVDV 4560
QY 130 ----- 129
Db 4561 PHEQVSINPDSIALIHGSEKLSVAQLDRESDRVARWLHRSPSSDTLIAVLAPRSCETII 4620
QY 130 ----- 129
Db 4621 AFLGILKANLAVPLDVKAPAARIDAIVSLPGNKLILLGANVTPPKLQEAADIFVPIRD 4680
QY 130 ----- 129
Db 4681 TPTTLTDGLQDQPTIERPSAOSLAYAMFTSGTRPKGMVQHRNIVRLVKNNSVNAKQ 4740
QY 130 ----- 129
Db 4741 PAAARIAHISNLAPDASSWEIYAPLLNGGAIVCADYFTTIDPQALQETFOEHEIRGAMLP 4800
QY 130 ----- 129
Db 4801 PSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQARLVGSGVFNAYGPTENTILSTIY 4860
QY 130 ----- 129
Db 4861 NVAENDSFVNGVPIGSAVNSGAYIMDKNQQLVPAGVMGELVVTGDLARGYMDPKLDAD 4920
QY 130 -----HTAKTITFRP----- 139
Db 4921 RFIQLTVNGSEQVRAVYRTGDRVYRPQFQIEFFGGRMDQOIKIRGHRIEPAEVEQAFIIND 4980
QY 140 ----- 139
Db 4981 GFVEDVAIVIRTPENQEPWAFVTAKGNSAREEBEATTQIEGWEAHFEGGAYANIEEIE 5040
QY 140 -----PPCGR----- 144
Db 5041 SEALGYDFMGWTSMYDGTBIDKDEMRWLNDTMRSLLDGKPAGRVLEVGTGTGMIMFNILG 5100
QY 145 ----- 144
Db 5101 RSQGLERYIGLEPAPSAABFVNNAKSPFGLAGRAEVHVGTAAADVGTLOGLTSDMAVINS 5160

QY 145 -----VFSM----- 148
Db 5161 VAQYPTPEYLAETIKSLVQVFCMKRIYLGDMRSMWAMNRDFAAARAAYSILADNASKDRVR 5220
QY 149 -----TCLSEML----- 155
Db 5221 QKMMELEKEEBELLVDPAFFTALASQLQDRIQHVEILPKRMKATNELSSRYAAVLHISD 5280
QY 156 -----NVSKRNDTGE----- 165
Db 5281 EPLPIYKIDPEAWINFEGRSLTREALAQVLKENENAEVAISNIPYSKTVVERHIVRSLD 5340
QY 166 ----- 165
Db 5341 QEDANAPESMDGSDWISAVRTRAQOCHTSLASDLFDIAEDAGFRVEVSWARQHSQHGAL 5400
QY 166 -----QG----- 167
Db 5401 DAVFHLKPAEDSRVLKFPPTDQORPLKSLTNQPLLPQAQSRRAELLIREGLQTLPPY 5460
QY 168 -----CGNFT----- 172
Db 5461 MIPSOITLIDRMLNANGKVDRELLARRAKITQSKPVEDIVPPRNSVEATVCKGTDVL 5520
QY 173 -----TFNPMFNV----- 181
Db 5521 GVEVGITDN--FFNLGGHSLMATKLAARLGRQLNTRISVRDVPDQPVVADLAAVIQNSA 5578
QY 182 ----- 181
Db 5579 PHEPIKPADYTCPVQSFQAQRLWFLDQLNVGATWYMLPLGIRLHGSLRVDALATAISAL 5638
QY 182 ----- 181
Db 5639 EQHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKDAYLAVLKHEQTTLPLDATE 5698
QY 182 PRWTKL----- 188
Db 5699 PGMWRVALLRLGEBEELHSIVMHIIISDGWSVEVLFDEMHRFYSSALRQDQPMELPLPI 5758
QY 189 ----- 188
Db 5759 QYRDPAAQKTEQVAEHQRLQDYWTEHLADSTPAELLTDLPPRPSILSGRANELPLTIEG 5818
QY 189 ----- 188
Db 5819 RLHDKLRAFCRVHQATPPFVILLALRAAHYRLTGAEDATLGTPIANRRPELENMIGFFV 5878
QY 189 ----- 188
Db 5879 NTQCMRIAIBENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQVI 5938
QY 189 ----- 188
Db 5939 LAVHSQQDLGKLTLEGLRDEAVDSAISTRPFVEFHLFEHADRLSGSVLYAKELFKURTIE 5998
QY 189 ----- 188
Db 5999 SVSVVFLETLREALDQPLTPLAVLPLTDGVEIGASKGLLDVPRTDYPRDANIVEVFQHV 6058
QY 189 ----- 188
Db 6059 RATPDAIAVKOATSILTYAQLDQOSDRLAIWLSRRHMPETLVGLAPRSCETIIMFGI 6118
QY 189 ----- 188
Db 6119 MKANLAYPLDINSAPAARLSILSAVDGNKVLVLSGSGVTAPQENPEVEAVGIEIAGT 6178
QY 189 -----YV----- 190
Db 6179 GLDKTQGSNARPSATSLAYVIFTSGTGKPGKGMVHRSVTRLAKPSNVISKLPOGARVA 6238
QY 191 ----- 190

Db	6239	HLANAFDASIWEIATTLNGLATVCLDYHTVLCRDLKEVERESITVTLMPALLKQC	6298	Db	7318	YGALQSFCTRSTVTTFVLLTVFRIAHFRLTAVDDATIGTPIANRNRPELETIVGCFVNT	7377
Qy	191	-----GPTKVNVDSTIY-----	203	Qy	223	-----	222
Db	6299	VAEIPETLAHLDLLYTGDRVGGHDAMRARSVLKIGMPSGYGPPTENTVIS-TIYEVDADE	6357	Db	7378	QCMRISIADDDNPEGLVRQVRNVATAAYANQDVPPERIVSALVPGSRNTRNPLVQLMFA	7437
Qy	204	-----	203	Qy	223	-----FYLIV-----	226
Db	6358	MFVNGVPIGKTVNSGAYVMDRNOQLVPSGVVGBLVYTGDLARGYTDPSLNKNRFYIIT	6417	Db	7438	VQSVEDYDQVRLEGLSVMMPEASTRFDMEFHLVPGDKLGTSGVLYSSDLFEQGTQINF	7497
Qy	204	-----	203	Qy	227	-----NMSR-----NLFR-----	235
Db	6418	VNGESIRAYRTGDRVYRPHDLQIEFFGRMDQVKIRGHRIEPEGESALLSHNSVQDAA	6477	Db	7498	VDIFQCLRSVLDQPLTPISVLPFSNAISLESLLDLEMTSDYPRDRTVVDLFRQAAI	7557
Qy	204	-----	203	Qy	236	-----	235
Db	6478	VVICAPADQSGABWAFVAARNTEDTQEBEAVDQGWETHFETAAYSEVKDIRQSE	6537	Db	7558	CPDSIAVKSSQLTYAQLDEQSDRVAWLHRRHMPAESLVGLSPRSCETIIAYFGIMK	7617
Qy	204	-----FLGLTAL-----	210	Qy	236	-----VPK-----	238
Db	6538	VGNDFMGWTSWYDGEIDKTMHEWLNMTMRMILDAREPGHVLIEGTGTGMWFMFLAKCP	6597	Db	7618	ANLAVILPDVYAPDARLAAILDVTSEGERLLLLGAGVQPQIGIQLRLSTAYIAEALSHATT	7677
Qy	211	-----	210	Qy	239	-----	238
Db	6598	GLQYVGFEPKSGAAQFVNDAAQSPALKDGRSIVHVTATDINKAGPIQPLRVVINSVA	6657	Db	7678	VDVTIPOPATSATLAVYIFTSGTGKPGVMIEHGRIVRLVRDINNVPFSGSALPVSH	7737
Qy	211	-----	210	Qy	239	-----	238
Db	6658	QYFPTPEYLFRVVEALVQIPSVVERIVFGDMRTNAINRDFVASRALHTLQKANKRLVRQM	6717	Db	7738	FSNLAWDAATWEIYTAVLNGTVVCIIDRDTMLDIAALNSTFRKENVRNAFFTPAFKQCL	7797
Qy	211	-----	210	Qy	239	-----	239
Db	6718	IYELEANEELTDPAFFTSLRTLGEKIKHVEILPKTMKATNELSKYRYAAVLHVRGSR	6777	Db	7798	AETPELVANLEILHTAGDRLDPGDANLAGTAKGGIFNVLGHTEHTAYSTFFPVVGEETFF	7857
Qy	211	-----	210	Qy	240	ING-----	242
Db	6778	EQSTIHQVSNAMIDFAADGLDROTLINLLKEHKDAGTVAIGNIPYKTIIVERFVNKSLS	6837	Db	7858	VNGVPGRGISNSHAYIIDRHQKLPAGVMGSELILITGDGVARGYTDSALNKDRFVYIDIN	7917
Qy	211	-----LLRYAQRNCT-----	220	Qy	243	-----	242
Db	6838	EDDMEEGONSLDGSAMVAARVMAAQSPSLDAMDVKEIAQAGYQVEVSWARQWONGAL	6897	Db	7918	GKSTWYRTGDKARYPRDQGLQEFFGRMDQWKIRGVRIEPEGEVELTLLDHKSVLAAATVV	7977
Qy	221	-----	220	Qy	243	-----	242
Db	6898	DAIFHFPEPKEGARTLIEPPTDYEGRNVTLTNRPLNSTQSRRLTQIREKLQTLPPY	6957	Db	8038	DKOLALRAQTVQKRSTAAARVPPRDEVEAVLCEBYSNLLVEVEGITDGFDFLGGHSLLA	8097
Qy	221	-----	220	Qy	243	TKLK-----NT-----	248
Db	6958	MIPSRIMVLDQMPVNNNGKIDRKELVRAIVAPKPRSAATRVAPRNEIEAILRDEFEDVL	7017	Db	8098	TKLAARLSQLNTRVSVDKDFDQPIADLADIIRRGSHRHDPIPATPYTGPVEQSPAQR	8157
Qy	221	-----HS-----	222	Qy	249	-----	248
Db	7018	GTEVSVLDNFFDLGGHSLMATKLAARVSRRLDAHISIKDVFQPVLDLAASIQRESAPH	7077	Db	8158	LWFLEQLNLGASWYLMFAIRMGRPLQTKALAVNALVALVHRHEALRTTFEDHDGVGVQVI	8217
Qy	223	-----	222	Qy	249	-----MRKLEKQKQ-----	257
Db	7078	EPIPORPYTGPAPQSPAQGRMLFMDQLNLGATWYLMPLAIRINGQLRVAALSAALFALER	7137	Db	8218	QPKSSQDLRIIDLSDAVDDTAYLAALKREQTAFDLTSEPGWRVSLRLRGDDDDYILSIWM	8277
Qy	223	-----	222	Qy	258	-----PVKEQ-----	262
Db	7138	RHETLTTFEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRRQTVFPDLSSEPG	7197	Db	8278	HHIISDGTWTVLRLQELQGFYSAAIRGEPLSQAKSLPIQYRDFAVMQRQENQIKEQAKQ	8337
Qy	223	-----	222	Qy	263	-----	262
Db	7198	WRVCLVKTGEEDHVLIVMHHIYDGVSDVILRGELQGFYSAAIRGEPLSQAKSLPIQYRDFAVMQRQENQIKEQAKQ	7257	Db	8338	LKYWSQQLADSTPCFEFLDLPRLPSILSGEADAVPMVIGDVTYQLLTDPCRTHQVTSFSLV	8397
Qy	223	-----	222	Qy	263	-----FEKKAKK-----	269
Db	7258	RDFAAWQREAKQVEEHQRQLGYWSKOLVDSTPAELLTDLPRLPSILSGRAGSVDDVTIEGVS	7317	Db	8398	LAAFRTAHLRTGTLDATVGTPIANRNRPELEGLIGFFVNTQCMRMAISETETESLSVQO	8457
Qy	223	-----	222				

QY	270	TOSTTTPFS-----	279	QY	304	-----	303
Db	8458	VRLTTTEAFANQDPPEQIVSTILLPGSRDTSRPLVQVMFALQSQODLGRIOLEGMTDEA	8517	Db	9538	ARRAQVAKRKAVSARVAPRNDTEIVLCBEYADILGTEVIGITDNFFDMGSHLMATKLA	9597
QY	280	-----	279	QY	304	-----	303
Db	8518	LETPLSTRLDLEVHLPQEVGKLSGSLYSTDLFEVETIRGIVDVFLIILRGLQEPQORL	8577	Db	9598	RLSRRLDTRVTYKVEFDPKVLADLAASIEQSGTTPHLPASSVYSGPVEQSYAQGRWLFLD	9657
QY	280	-----	279	QY	304	-----	303
Db	8578	MAMPITDGITKLARDQGLLTVAKPAYPRESSVIDLFRQVAAAPDAFVWDSSLTLYADL	8637	Db	9658	QFNLNATWTHMSLAMRLGLPLNMDALDVALRALEQRHETLRTTFFBAQKDIGVQVVHEAGM	9717
QY	280	-----	279	QY	304	-----	303
Db	8638	DGQSNLAHLWCORNMAPETLVAFPAPRSLCTIVAFGLVKANLAYLPDVPNAPARIAE	8697	Db	9718	KRLKVLDSLCKNEKEHMAVLENEQMRPPTLASEPGWKHGLARLGTEYILSLVMMHMFSD	9777
QY	280	-----	279	QY	304	-----	303
Db	8698	IILSAVEGHKLVQAHGPELGLTMAOTELVQIDEALASSSGDHEQIHASGPTATSLAYV	8757	Db	9778	GWSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFPAWQKEAAQVAEHERQLAYWEN	9837
QY	280	-----	279	QY	304	-----	303
Db	8758	METSGTGKPGVMIDHRSIIRLVKNSDVVATLPTPVRMANVSNLAFDISVQEIYTALIN	8817	Db	9838	QLADSTPGELLTDFPRPQFLSGKAGVIVPTIEGPPVEKLLKFSKERQVTLFVSLLTAPRA	9897
QY	280	-----	279	QY	304	-----	STST 307
Db	8818	CGTLVCLDYLTLLDSKILYNVFEAQVNAAMFTVLLKQCLGNMPAIIISRLSVLFNVGDR	8877	Db	9898	THFRLTGAEDATIGTPIANRNRPELEHIIGFFVNTQCMRLLLDTGSTPESLVQHVRSVAT	9957
QY	280	-----	279	QY	308	IAY-----	310
Db	8878	LDAHDAVASGLIQDAVYNAYGPTENGMOSTMVYKVDVNEPFGVNGVPIGRSITNSGAYMD	8937	Db	9958	DAYSNQDIPFERIVSALLPGSRDASRPLIQMPALHSQPDIGNITLEGLEHERLPTSVA	10017
QY	280	-----	284	QY	311	-----	310
Db	8938	GNQQLVSPGVMEIVVTGDLARGYTDLSALDEDRFVHVITDGEENIKAYRTGDRVRYRK	8997	Db	10018	TFDMEFHLFQEPNKLSGSILFADELFPQETINSVVTVFQETLRRGLDQPVSIPTMLT	10077
QY	285	-----	284	QY	311	-----	310
Db	8998	DPEIEBFGMDQVKIRGRIEPAVEHALLGHDLVHDAAVLVRKPAQEPSEMIAPITQ	9057	Db	10078	DGLIDLEKGLLEIBSSNPPRDYSVVDVFRQVAAANPAPAVVDSETSMTSLDQKSEQ	10137
QY	285	-----	284	QY	311	-----	314
Db	9058	EDETIEQESNKVQVGEHFDVSRYADIKOLDTSTFGHDFLGWTSMDYGVDPVNMKE	9117	Db	10138	IAWLHAQGLRPESLICVWAPRSFETIVSLFGILKAGYAYLPDVSNSPAARIQPIILSEVE	10197
QY	285	-----	284	QY	315	-----	314
Db	9118	WLDETTASLLDNRPQGHILEIGAGTGMIILSNLKGVDGLQKYVGLDPAPSAALFVNEAVKS	9177	Db	10198	GKRLVLLSGIDMPSQDRMDVETARIQDILITNTKVERSDPMSRPSATSLAYVIFTSGSTG	10257
QY	285	-----	284	QY	315	-----	314
Db	9178	LPSLAGKARVLGTALDIGSLDKNEIQPELAVINSVAQYFPTSEYLIKVVKAVEVPSVK	9237	Db	10258	RPKGVNIEHRNILRLVKQSNVTSQLPQDLRMAHISNLAFAASIWEIFTAILNGGALLCID	10317
QY	285	-----	287	QY	315	-----	318
Db	9238	RVFFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELESEBELVDPAFFVSLRS	9297	Db	10318	YFTLDSQALRTTFEKARVNATLFPALLKECLNAPTLPEDLVKLYIGGDRDATDAK	10377
QY	288	-----	287	QY	319	-----	318
Db	9298	QLPNIKHVEVLPKMKATNELSSRYAAVLHISHNEEBQLLIQDIPTAWVDAATQKDS	9357	Db	10378	IQALVKTIVYNAYGPTENTVMSTIYRLTDGESYANGVPIGNVSSGAVIMDKQRLVPP	10437
QY	288	-----	303	QY	319	-----	318
Db	9358	QGLRNLQOGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAIC	9417	Db	10438	GVWGLVVGDLGARGYTNSTLNADRFVDIVINDQKARAYRTGDRTRYPKDGSIIEFFGR	10497
QY	304	-----	303	QY	319	-----	318
Db	9418	TSFDAPALQLAKEGFRVELSWARQSONGALDAVFHRLATDANCERSRVLPHPTDQ	9477	Db	10498	MDQQVKIRGHRVEPAEVBQAMLGNKAIHDAAVVQAVDQETEMIGFVMSADRSEGE	10557
QY	304	-----	303	QY	319	-----	318
Db	9478	GRQLRTLNRPLQRAQSRRIESQVFEALQATALPAYMIPSRIIIVLPQMPNTNANGKVDKOL	9537	Db	10558	EITNOQWEDHFESTAYAGIEAIDQATLGRDFTSWTSMYNGNLIDKAEMEELDDTMQS	10617
Db				QY	319	-----	318

Db	10618	LLDKEDARPCAEIGTGTGMVLENLPKNDGLESYVGIIEPSRSALFVDKAAQDPFGLQKGT	10677	Db	11698	QAVKLHANSVRFRTRISDALVESGSPTEBELSTRPTAQSLAYVMFTSGSTGPKGVNVEHR	11757
Qy	319	-----	318	Qy	346	-----	345
Db	10678	QILVGTABDIKLVKDFHPDWWVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW	10737	Db	11758	GITRLVKNNSVNAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGTVCIDYTTIDIKAL	11817
Qy	319	-----	318	Qy	346	-----	345
Db	10738	ATNRDFLVSRALYTLGDKATKDQIRQEVARLEENEDELLVDPAFFTSLSQWPKVKHVE	10797	Db	11818	EAVFKQHHIRGAMLPALLKQCLVSAPTMISSLEILFAAGDRLLSSQDAILARRAVSGVY	11877
Qy	319	-----	318	Qy	346	-----	345
Db	10798	ILPKRMETSNESSRYAAVLHICRDGEGNRNRYGRVHSVEENAWIDFASSGMDRHVLQ	10857	Db	11878	NAYGPTENTVLSTIHNIGENAEFSGVPICGNAVSNSGAFVMDQNOQLVSAGVIGELVVVG	11937
Qy	319	-----	318	Qy	346	-----	345
Db	10858	MLDERDAKTAIGNIPHSNTINERHFTTSLDTEGEGIAQDSLDGSAMOSATKAMAARCP	10917	Db	11938	DGLARGYDTSKLVRDRFIYITLDGNRVRAYRTGDRVRHRPKDQIEFFGMDQOIIRGH	11997
Qy	319	-----	318	Qy	346	-----	345
Db	10918	CLSVTELVEIGQAAGPRVEVSWARQSRQHGALDVVPHHLEDDRGRVLINFTDFELPP	10977	Db	11998	RIEPAEVEQALARDPAISDSAVITQLTDEEPELVAFPSLKGNGANGTNGVNGVSDQEKID	12057
Qy	319	-----	318	Qy	346	-----	353
Db	10978	STGLTSRPLQRIQNRFPESQIREQLOTLPPYVMPVSRIVVLERMPLNANSKVDRKELARK	11037	Db	12058	GDEQHALLMENKIRHNLAQLLPTMIPSRIIHVDQLPVNANGKIDRNLAVRAQAATPTS	12117
Qy	319	-----	325	Qy	354	-----	356
Db	11038	ARTLQTIKPSATRVAPNDIEAVLCDEFQAVLGVTVGVMDNPFELGGHSLMATAARLS	11097	Db	12118	SVSTVAPNDIETIICKEFADILSVRGITDNFDFLGGHSLIATKLAARLRRLDTRVS	12177
Qy	326	-----	325	Qy	357	-----	356
Db	11098	RRLDTRVSVKDIENQPILOLADVQVQTSAPHEAIPSTPYSGPVEQSFQGRMLFIDLQNL	11157	Db	12178	VRDVFDPVVGQLAASIQQGSTPHEAIPALSHSGPVQSQFAQGRMLFDRFNLAANYTM	12237
Qy	326	-----	325	Qy	357	-----	365
Db	11158	LNASWYHMLASRLRGLRIEALQSALATIEARHESLRTTFEQDGPVQVIVRAARNKQL	11217	Db	12238	PFVRLRGLRPLVDALQTLALRALEERHELLRTTFEQDQVGMQIVHSRMRDICVVDISGA	12297
Qy	326	-----	325	Qy	366	NE-	367
Db	11218	RIIDVSGTEDAYLAALKQEQDAAFDLTAEPGRVALLRGLDPDHHVLSIVMHHIISDGWSV	11277	Db	12298	NEDLAKLEEQQAPFNLSDEVAVRWALFKAGENHHILSIVMHHIISDGWSVDIFQOELAQ	12357
Qy	326	-----	331	Qy	368	-----	367
Db	11278	DILRQELGQLYSNASSQAPLPPIQYRDFALWQDSQIAEHQKQVYWKQVLSNKPSEL	11337	Db	12358	FYSVAVRGHDPLSVQVKPLPIHYRDFAVMQRDQVAVHESQLQYWIEQLADSTPAEILSD	12417
Qy	332	-----	339	Qy	368	-----	367
Db	11338	LADFTPKALSGDADVIPIEIDDQVYONLRSFCRARHVTSFVALLAFAAHYRLTGABD	11397	Db	12418	FNRPEVLSGEAGTVPVIVIEDEVYEKLSLFCRNHQVTSFVLLAFAFVAHYRLTGADATI	12477
Qy	340	-----	339	Qy	368	-----	367
Db	11398	ATIGSPIANNRPEGLIGCFVNTQCLRIPVKSEDTFDTLVKQARETATEAQDNQDVFF	11457	Db	12478	GTPIANRNRPELDELIGFFVNTQCMRIALAEHDNPLSVVRRVRSSTAASAFENQDVFFERL	12537
Qy	340	-----	343	Qy	368	-----	371
Db	11458	ERIVSSWASSRDTSRNPLVQVMFAVSHQDLGNIRLEGVEGPKVSMASATRFDAEMHLF	11517	Db	12538	VSALLPGSRDASRNPLVQVMFVHVSQRNLGKLQLEGEPTPYTATTTFDVFPHLEFD	12597
Qy	344	-----	343	Qy	372	-----	371
Db	11518	EDQGM/LGNNVFSKOLFSEETIRSVVAVFOETLRRGLANPHANLATPLTDGLPSLSLC	11577	Db	12598	KSLAGNVVFAADLFEAATIRSVVEVFHEILRRGLDQPDIAISTMPLVDGLAALNSRNLP	12657
Qy	344	-----	343	Qy	372	-----	377
Db	11578	LQVNPDPYRDASVIDVPREQVASIPKSTAVIDASSQLTYTTELDRSSQLATWLRQVTV	11637	Db	12658	VEDIEPDFATEASVVDVFQTVVNPDALAVTDTSTKLTAYELDQSDHVAWLSKQKLP	12717
Qy	344	-----	345	Qy	378	-----	377
Db	11638	PBELVGLAPRSCETIIAFGLIIGKANLAYPLDVNAPAGRIETILSSLPNRLILGSDT	11697	Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMDSNVPEARQAILSEIPGKFKVLLGAGV	12777
Qy	346	-----	345	Qy	378	-----	377
				Db	12778	PIDPNKTADVRMVFISDIVASKTDKYSYSGTRPSASSLAYVIFTSGSTGRPKGVNVEHRG	12837

QY 378 ----- 377
Db 12838 VISLVKQNASRIPQSLRMAHVSNLAFDASVWEIFTTLLNGTILFCISYFTVLDSKALSA 12897
QY 378 -----MSSLY-----YN-- 384
Db 12898 FSDHRINITLLPPALLKQCLADAPSVLSLESYIGGDRLDGADATKVKDLVKGKAYNAY 12957
QY 385 ----- 384
Db 12958 GPTENSVMSTIYTIETHPANGVPIGTSGLPKSKAYIMDQDQLVPAGVMGELVAGDGL 13017
QY 385 -----E 385
Db 13018 ARGYTDPSLMTGRFIHITIDGKQVAYRTGDRVRYPRDYQIEFFGRLDQOIKIRGHRIE 13077
QY 386 TMEVENKTADS----- 397
Db 13078 PAEVEQALLSDSSINDAVVSAQNKEGLEVMGVITTOAQSVDKEASNKVQWEAHFDS 13137
QY 398 ----- 397
Db 13138 TAYANIGIDRDALGQDFLSWTSWYDGLIPREMQEWLNDTWRSLLDNQPPGKVLEIGT 13197
QY 398 -----NKTTPSPM----- 407
Db 13198 GTGMVLFNLGKVEGLQSYAGLEPSPSVTAVWVKAIBTFPSLAGSARVHVGTAEIDISSDG 13257
QY 408 -----GFQRTF----- 413
Db 13258 LRSDLVINSVAQYFSPREYLAELTANLIRLPGVKRIFFGDMRTYATNKDFLVARAVHTL 13317
QY 414 ----- 413
Db 13318 GSNASKAMVRQVAKLEDDEBELLVDPAPFTSLSDQPPDEIKVEILPKRMAATNELSSY 13377
QY 414 ----- 413
Db 13378 RYAAVHVGHGHPNGEDEDKQAVKDINPKAWVDFAGTMDRQALLQLLQDRQRDDVV 13437
QY 414 ----- 413
Db 13438 AVSNIPYSKTIEMRHLSQLDDEDETSADVGTAMISRTQSRACEPALSADLIBIGK 13497
QY 414 ----- 413
Db 13498 IGFEVEASWARQHSORGGILDVPHRFEPPRPHSHVMFRPTEHKGRSSSLTNRPULILQ 13557
QY 414 ----- 413
Db 13558 SRRLEAKVRERLQSLPPYMPISRITLLDQMLTSGKVDKRLARQARVIPSAASTLD 13617
QY 414 ----- 413
Db 13618 FVAPRTEIEVLCEEFTDLLGVKVGITDNFFELGHSLLATKLSARLSRLDAGITVKVQ 13677
QY 414 -----IDPLM----- 418
Db 13678 FDQPVLAADLAASILQSSRHSRISPSLPEGPVEQSPAQRNLWFLDQFNIDALWYLPFAL 13737
QY 419 ----- 418
Db 13738 RMGPQLQVDAALAALEERHESLTTFFERDGVGIQVVQPLRTTKDIRIIDVSGMRDD 13797
QY 419 ----- 418
Db 13798 DAYLEPQEQOTPFPLASEPGRVALLKLGKODHILSIWMHIIISDGNSTEVLORELQ 13857
QY 419 -----D 419
Db 13858 FYLAAGSGKAPUSQVAPLPIQVRDFAVWQREBQVAESQRLDYWKQLADSPAEILLAD 13917

QY 420 Y----- 420
Db 13918 YTRPNVLSGEAGSVSFVINDSVYKSLVFCRQRQVTTFTTLAFAAHYMTGSDDATI 13977
QY 421 ----- 420
Db 13978 GTPIANRPELENLIGCFWNTQCMRITIGDDTFESLVQOVRSTTATAFENQDVPPERI 14037
QY 421 -----LDSLLF----- 426
Db 14038 VSTLSAGSRDTSRNPLVQLLFAVHSQQGLRIQLDGVWDEPVLSTVSTRFDLSFHAQEA 14097
QY 427 -----LDEI----- 430
Db 14098 DRLNGSMVFATDLFQPETIQGFVAVVEEVLQRLGEQFQSPMATPLAEGIAQLRDAGALQ 14157
QY 431 -----RNFSL----- 435
Db 14158 MPKSDYPRNASLVDFVQOQAMASPSVAVTVDSTSKUTYAEGLRSLDQAAASYLRRQQLPAE 14217
QY 436 -----RS----- 437
Db 14218 TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEAIISVPGRRLLVSGVRH 14277
QY 438 ----- 437
Db 14278 ADINVPNAKTMLTISDTVTGTAIGTPEPLVVRPSATSLAYVIFTSGTGKPKGMVEHRA 14337
QY 438 ----- 437
Db 14338 IMRLVKDSNVTHMPATRMAHVNTIAFVSLUFEMCATLLNGTLVCIDYLTLLDSTMRL 14397
QY 438 ----- 437
Db 14398 ETFEREQVPAALFPPALLRQCLVNMPPDAIGMLEAVYVAGDRPHSRDARATQALAGPRVYN 14457
QY 438 -----PT-----YVN-----LTPP----- 446
Db 14458 AYGPTENAILSTIYNIDKHDYVNGVPIGSAVNSGAYVMDRNQQLLPQVGMELVVTGE 14517
QY 447 ----- 446
Db 14518 GVARGYTDASLTDTRFVTVIDGQRQRAYRTGDRVRYRPGFQIEFFGRLDQOAKIRGHR 14577
QY 447 -----EH----- 448
Db 14578 VELGEVEHALLSENSVTDAAVLRMTMEEDPOLVAFVTTDHEVYRSGSSNEEDPYATQAA 14637
QY 449 -----RAVNLSSTNS----- 459
Db 14638 GDMRKRLRSLPPYMYVPSRVITLRQMLNANGVKDRKLARRAQMTPPTASSSGPVHVAPR 14697
QY 460 ----- 459
Db 14698 NETEAAICDEFITILGVKVGITDNFFELGHSLLATKLAARLSRMLRISVKDLFDDPV 14757
QY 460 ----- 459
Db 14758 PVSLLAKLEQQQGPSGEDESSTVGIVPFQLLPAEMSREIIQRDVVPOIENGHSTPLDMYP 14817
QY 460 ----- 459
Db 14818 ATOTQIFFLHDKATGHATPPPLSLDFPETADCRRLASACAALVQHFDIFRTVFSRGGR 14877
QY 460 ----- 459
Db 14878 FYQVVLHLDVPEVETEQEDELVALHEADKQOPLRGLRAMLRILKRPKAKORLV 14937
QY 460 -----LW----- 461
Db 14938 LRMSHSLYDGLSLEHIVNALHALYSDKHILAQAPKFGLYMHMHASRRAGYNFWRSILOGS 14997
QY 462 -----W----- 462

Db	14998	SMTSLKRSVGALEAMTPSAGTWTQTSKSIIRIPPAALKNGITQATLFTAAVSLLAKHTKST	15057
QY	463	-----	462
Db	15058	DVVFGRVSRQDLSINCQDIVGPCINEVPRVRIDEGDDMGLLRAIQDYTSRPHET	15117
QY	463	-----W 463	
Db	15118	LGLQEVKENCTDW	15130
RESULT 9			
ADP84155	ID ADP84155 standard; protein; 22157 AA.		
XX	AC ADP84155;		
XX	DT 26-AUG-2004 (first entry)		
XX	Human CA125 protein sequence SeqID 5.		
DE	human; CA125; antigenic determinant; ovarian carcinoma cell;		
KW	serum tumour marker; ovarian cancer; cytostatic.		
XX	Homo sapiens.		
OS	Key Location/Qualifiers		
FT	Misc-difference 1. :22152		
FT	/label= Xaa		
FT	/notes= "Xaa can be any naturally occurring amino acid"		
XX	WO2004045553-A2.		
PN	03-JUN-2004.		
XX	17-NOV-2003; 2003WO-US037041.		
PF	15-NOV-2002; 2002US-0427045P.		
XX	(UYAR-) UNIV ARKANSAS.		
PA	O'brien T, Beard J, Underwood L;		
PI	WPI; 2004-420533/39.		
DR	N-PSDB; ADP84151, ADP84152, ADP84153, ADP84154.		
XX	New isolated nucleic acid molecule encoding CA125 protein, useful for		
PT	diagnosing or treating cancer, particularly ovarian carcinoma.		
FT	Claim 14; SEQ ID NO 5; 501pp; English.		
XX	This invention relates to a novel nucleic acid molecule identified as		
CC	CA125, and the encoded protein thereof. Specifically, it refers to the		
CC	cloning, identification and expression of the CA125 gene, and in		
CC	particular the glycosylated amino terminal domain, multiple repeat domain		
CC	and the carboxy terminal domain. The present invention describes CA125 as		
CC	an antigenic determinant that is expressed on ovarian carcinoma cells,		
CC	yet show essentially no expression in normal adult ovarian tissue.		
CC	Accordingly, this protein represents a serum tumour marker that can be		
CC	used for early diagnosis and improved prognosis in cancer patients,		
CC	specifically ovarian cancer, such that they can be used to develop		
CC	pharmaceutical compositions that exhibit cytostatic activities. This		
CC	polypeptide sequence is the human CA125 protein of the invention.		
XX	Sequence 22157 AA;		
SQ	Query Match 22.9%; Score 561; DB 8; Length 22157;		
	Best Local Similarity 1.2%; Pred. No. 0.0027;		
	Matches 249; Conservative 77; Mismatches 137; Indels 20726; Gaps 80;		
QY	2 GRKEMVRDVPKMFVLIS-----		19

Db	755	GTPEISTTKPSSAVLSSMTLSNAATSPERVRNATSPLTHPSPSGEETAGSVLTLSAE	814
QY	20	-----	19
Db	815	TTDSPNIHPTGTLTSESSPSTLSLPVSGVKVTFSSSTPSTHLPSTSGEETEETSNPSV	874
QY	20	-----	19
Db	875	SQPETSVSRVRTTLASTSVPTPVFPTMDTWPTRSAQFSSSHLVSELRATSTSVTNSTGS	934
QY	20	-----	19
Db	935	ALPKISHLTGATMSTQNRDTFNDAAPOSTTWTPETSPRFTGLPSATTVTSTATLSA	994
QY	20	-----	19
Db	995	TVMSKFTSPATYSMEATSIREPSTTILTTETNGPGSMASVASTNIPICKGYITEGRDLT	1054
QY	20	-----	19
Db	1055	SHLPIGTTASSETSMDFTWAKESVMSVSPSQSMDAAGSSTPGRTSQFVDTPFDDVVHLT	1114
QY	20	-----	19
Db	1115	SREITIPRDGTSSALTPQMTATHPPSPDPGSGARSTWLGLSSSPSSPTPKVTMSSTFSTQ	1174
QY	20	-----	19
Db	1175	RVTTSIMDVTVTSRWNPENLFTSLTSPSNIPTSGAIGKSTLVPLDTPSPATSEASEG	1234
QY	20	-----	19
Db	1235	GLPTLSTYPESTNTPSIHLGAHASSESPSTINLTMAVVVKPGSYTLTPFSIETHIVST	1294
QY	20	-----	19
Db	1295	ARMAYSSGSSPENTAPGETNCGTWDPPTYITTTDPKOTSSAQVSTPHSVRTLRTTENHP	1354
QY	20	-----	19
Db	1355	KTESATPAAYSGSPKISSPNLTSPATKAWTITDTTEHSTQLHYTKLAESGFEQSQAP	1414
QY	20	-----	19
Db	1415	GPVSVVIPTSPITGSGTLELTSVDPGEPLVLAPSEQTTITLPMATWLSTLTBEMASTDL	1474
QY	20	-----	19
Db	1475	DISSPSPMSTFAIPFPMSTPSHELKSEADTSAIRNTDSTTLDQHLGIRSLGRTGDLTT	1534
QY	20	-----	19
Db	1535	VPITPLTTTWSVIEHSTQADTSLSATMSPHVTQSLKDOTSIPASAPSHLTVVPELG	1594
QY	20	-----	19
Db	1595	TQGRSSSEATFEWKPSTDTLSREIETGPTNIQSTPMDNTTTGSSSSGVTLGIAHLPIGT	1654
QY	20	-----	19
Db	1655	SSPAETSTNNALERRSSTATVSMAGTMGLLVTSAPGRSISQSLGRVSSVLSESTTEGVTD	1714
QY	20	-----ISPL-----	23
Db	1715	SSKGSPLNTQGNLTALSSSLEPSVABSQMSTSIPLTSSPTTDPVEFIGGSTFWTKVT	1774
QY	24	-----	23
Db	1775	TWMTSDISKSSARTESSATLMSTALGSTENTCKEKLRTASMDLPSPTPSMEVTPWISLT	1834
QY	24	-----	23
Db	1835	LSNAPNTTDSLDSHGVTSSAGTLATDRSLNTGVTTRASRLENGSDTSKSLSMGNSTHT	1894

QY	24	-----	38
Db	1895	SMTDEKSEVSSIHPRPETSAPGAETTLTSTPGNRAISLTLPLFPSSIPVEEVISTGITSG	1954
QY	24	-----	38
Db	1955	PDINSAPMTHSPITPTTIWVTSTGTIEQSTOPLHAVSEKVSQVOSTPYVNSVAVASP	2014
QY	24	-----	38
Db	2015	THENSVSSGSTSTSPYSSASLESLSLDSISRRAITSLWLDLTLTSLPTTWPTSLSEALS	2074
QY	24	-----	38
Db	2075	SGHSGVSNPSTTTTEPPLFSAASTSAKORNPETETHGPQNTAASTLNTDASSVTGLSET	2134
QY	24	-----	38
Db	2135	PVGASISSEVPLPMAITSRSDVSGLTSESTANPSLGTASAGTKLRTTISLPTSESLVSF	2194
QY	28	-----	38
Db	2195	RMNKDPWTVSIPLGSHPTTNTETSIIPVNSAGPGLSTVASDVLDTPSDGARSIPTVSFP	2254
QY	28	-----	38
Db	2255	SPDTEVTTISHPEKTHSFRITSSILTHELTSRVTRIPGDMSSAMSTKPTGASPSITLG	2314
QY	28	-----	38
Db	2315	ERRTISAAPTSPIVLTASTETSTVSLDNETTVKSIDLDARKTNELPSDSSSDLI	2374
QY	29	NCKVM-----	33
Db	2375	NTSIASSTMDVTKTASISPTISGWTASSPSLFSDDRPOQVPTSTTETNTATSPSVSNT	2434
QY	34	-----	38
Db	2435	YSLDGGSNVGGTPSTLPPFTIHPVETSSALLAWSRPVFTFTWVSTDTDTSNGENPTSSNS	2494
QY	34	-----	38
Db	2495	VVTSVPAGTWASVGGTTDLPMAGFLKTPAGEAHSLLASTIEPATAPAFPHLSAAVVVGS	2554
QY	34	-----	38
Db	2555	SATSEASLLTTSESKAHSSPQPTPTTSGANWETSATPESLLVVTETSDTLTSLKILT	2614
QY	39	-----	38
Db	2615	DTILFSTVSTPPSKFPSTGTLGASFPILLPDTPAIPLTATEPTSSLATSPDSTPLAVTIA	2674
QY	39	-----	38
Db	2675	SDSLGTVPTTLTMTSETNGDALVLKTVSNPDRSIPGITIQGVTESPLHPSSTSPSKIYA	2734
QY	39	-----	38
Db	2735	PRNTVEGSI VALSTLPAGTTGSLVFSQSSSENSETTALVDSSAGLERASVMPLTGSGQ	2794
QY	39	-----	38
Db	2795	MASSGGIRSGTHSTGKTTFSSLPLTMNPGEVTAMSEITNRLTATQSTAPKGI PVKPTS	2854
QY	39	-----	38
Db	2855	ABSGLLTPVSASSSPSKAFASLTAPPSTWGIPOSTLTFFSEVPPLDTSASLPTPGQS	2914
QY	39	-----	38
Db	2915	LNTIPDSDASTASSLSKSPKPNRPRMMTSTKAISSASSFQSTGTTETPEGSASPMSMAGH	2974
QY	39	-----	38
Db	2975	EBRVPTSGTGDPRYASESMSYPDPKASSAMTSTSLASKLTLTFLSTGQAARSGSSSPIS	3034
QY	39	-----	38
Db	3035	LSTEKETSFLSPTASTSRKTSILPLGPMARQPNILVHLQTSALTSLTSTLNMSQEBEPPE	3094
QY	39	-----	38
Db	3095	LTSSQTIABEEGTAAETQTLTFTPTSETPTSLPVSSPTEPTARRKSSPETWASSISVPAK	3154
QY	39	-----	38
Db	3155	TSLVETTDGTLVTTIKMSSQAAQGNSTWPAABETGTSAGTSPGSPGVSTTLKIMSSKE	3214
QY	39	-----	43
Db	3215	PSISPEIRSTVNSPWKTPETTVPMETTVLQSTALGSGSTSISHLPTGTTSTKSP	3274
QY	44	-----	43
Db	3275	TENMLATERVSLSPSPPEAWTNLYSGTPOGTRQSLATMSSVSLESPTARSITGTGQSSP	3334
QY	44	-----	43
Db	3335	ELVSKTTGMEFSGMHGSGTGGTTHVLSLSTSSNILEDPTSPNSVSSLTDKSKHKETW	3394
QY	44	-----	43
Db	3395	VSTTAIPSTVLNKKIMAAEQTSRVSDEAYSGTSSWSDQTSDDITLGASPDVNTLYIT	3454
QY	44	-----	43
Db	3455	STAQTSLVSLPSGQGITSLNPSGGKTSSASSVTPSIGLETLRANVSAVKSDIAPTA	3514
QY	44	-----	43
Db	3515	GHLSTQSSPAEVSILDVTTAPTFGISITITMTGNSISTTTPNPEVGMSTMDSTPATER	3574
QY	44	-----	43
Db	3575	TTSTHPTWSTASASDSWTVTDMTNSLKVARSPTISTMTTSPFLASSTELDSMTPHG	3634
QY	44	-----	43
Db	3635	RITVIGTSLVTPSSDASAVKTETSTSERLSPSDTTASTPISFTRVQRMSSVDPILST	3694
QY	44	-----	43
Db	3695	SWTPSSTEADVPVSMVPTDHAASKTDPNTPLSTFLFDSLSTLDMTGRSLSSATATSA	3754
QY	44	-----	43
Db	3755	POGATTPQBLTLETWISPATSQLPFSIGHITSAVTPAAMARSSGVTFSPDPTSKAEQT	3814
QY	44	-----	43
Db	3815	STQLPTTTSAPGQVPRSAATLVDIPIHTAKTPDATFQOGQTALTTEARATSDSWNEKE	3874
QY	44	-----	43
Db	3875	KSTPSAPWITEMMNSVSEDTIKEVTSSSVLKDPEYAGHKLGIWDDFI PKFGKAAHREL	3934
QY	44	-----	43
Db	3935	PLLSPPQDKEAIHPSTNTVETTTGWVTSSEHSHSTIPAHASSKLTSPVVTSTREQAIV	3994
QY	44	-----	43
Db	3995	SMSTTTPESTARTEPNSFLTIELRDVSPYMDTSTTQTSIISSPGSTAIYKGRTEIT	4054
QY	44	-----	43

Db	4055	SYKRISSEFLAQMRRSDSPSEAITRLSNFPAMTESGGMILAMQTSPPGATSISAPTLD	4114	Db	5135	SGHESHSPALADSETPKATTQWVITTTVGDPAPTSMPPVHGSSETNIKREPTYFLTPRL	5194
QY	44	-----	43	QY	47	-----LSKI-----	50
Db	4115	SATASWTGTLATTQRTYSEKTTLPKGREDTQSPPCVBEETSSSSVVPIHATTSPS	4174	Db	5195	RETSTSQSSRPPTDTSFLLSKVPTGTITEVSTGTVISSSKISTDPHDKSTVPDPDTFTGEI	5254
QY	44	-----	43	QY	51	-----	50
Db	4175	NILLTSQGHSPSTPPTVSVFLSETSGLGKTTDMSRI SLEPGTSLPPNLSSTAGBALSTY	4234	Db	5255	PRVFTSSIKTKSAEMTITTOASPPESASHSTLPLDTSTLTLSQGGTHSTVSOQFPYSEVTT	5314
QY	44	-----	43	QY	51	-----	50
Db	4235	EASRDTKAIHHSADTAVTNMEATSEYSPIPGHTKPSKATSPLVTHIMGDITSSTVFG	4294	Db	5315	LMGMGPGNVSMWMTTPVEETSSVSLMSSPAMTSPSPVSSTSPOQIPSSPLPVLTALPTSV	5374
QY	44	-----	43	QY	51	-----	50
Db	4295	SSETTEIETVSSVNOQLQERSTQVASSATETSTVITHVSSGDATHTVTKQATPSSGTS	4354	Db	5375	LVTITDVLGTTSPESVTSSPPNLSI THERPATYKDTAHTAAMHHSNTWTAVTNVGTSGS	5434
QY	44	-----	43	QY	51	-----	50
Db	4355	ISSPHQFITSTNTFTDVSTNPSTSLIMTESSGVITITQTGPTGAATQGPYLLDTSTMPYL	4414	Db	5435	GHSQSSVLADSETSKATPLMSTASTLGDTSVSTPNISQTNQIQTEPTASLSPRLRES	5494
QY	44	-----	43	QY	51	-----	50
Db	4415	TETPLAVTDFMQSEKTTLISKPKDVTWTSPSVAETSYPSLTPFLVTTIPPATSTLQ	4474	Db	5495	STSEKTSSTTETNTAFSVVPTGAI TQASRTEISSRSTISOLDRSTIADPDISTGMITRLF	5554
QY	44	-----	43	QY	51	-----	50
Db	4475	GQHTSSPVSATSVLTSGLVKTTDMLTSMPEVTNSPONLNPSEILATLAATTDIETIH	4534	Db	5555	TSPIMTKSAEMTVTTQTTTPGATSQGILPMDTSTTLFOGGTHSTVSOQPPHSEITTLRSR	5614
QY	44	-----	43	QY	51	-----	50
Db	4535	PSINKAVTNMGTASSAHVLHSTLPVSSEPTSTATSPMVPASSMGDALASISIPGSETTOIE	4594	Db	5615	TPGDVSMWMTTPVEETSSGFSLMSPMTSPSPVSSTSPEISPSPLPVLTALLTSLVLTVT	5674
QY	44	-----	43	QY	51	-----	50
Db	4595	GEPTSSLTAGKENSTLOEMNSTESNII LSNVSGAITEATKMEVPSDATFIPTPAQS	4654	Db	5675	NVLGTTSPPEVTSPPNLSSTQERLTTYKDTAHTAMHASMHTNTAVANVGTSISGHS	5734
QY	44	-----	43	QY	51	-----	50
Db	4655	TKFPDIFSVASSRLNSPPMTISTHMTTQTGSSGATSKIPLALDTSTLETAGTPSVVT	4714	Db	5735	QSSVPADSHTSKATSPMGITFAMGDTSVVTS TPAFFETRIQSESTSSLIPLGLDRTTSEE	5794
QY	44	-----	43	QY	51	-----	50
Db	4715	EGFAHSKITTAMNDVKDVSQTNPFPQDEASSPSQAPVLVTTLPSSVAFTPQWHSTSSP	4774	Db	5795	INTVTETSVLSEVPTTTTTEVSRTEVITSSRTTISGPDHSMSPYISTETITRLSTFPF	5854
QY	44	-----	43	QY	51	-----	50
Db	4775	VSMSSVLTSSLVKTAGKVDTSLETVTS PQSMSNTLDDISVTSAAATDIEHTHPSINTVV	4834	Db	5855	VTGSTEMAITNQTGPIGTISQATLTLDTSSTASWEGTHSPVTQRPHESETTMSRSTKG	5914
QY	44	-----	43	QY	51	-----	50
Db	4835	TNVGTTGSAFESHSTVSAPEPSKVTS PNVTTSTMEDTII SRSEIPKSKTTRTETETSS	4894	Db	5915	VSWQSPPSVEETSSPSPVPLPAITSHSLSYSAVSGSSPTSALPVTSLTSGRRKKTIDML	5974
QY	44	-----	43	QY	51	-----	50
Db	4895	LTPKLRETSISOEITSSSTETSTVPYKELTGATTEVSRKTDVTSSTSPFPQDQSTVSLDI	4954	Db	5975	DTHSELVTSLSLPASSFSGEILTSEASTNTETIHFSENTAETNMGTNSMHLHSSVSIH	6034
QY	44	-----	43	QY	51	-----	50
Db	4955	STETNRLTSPIMTESABITITQTGPHGATSQDFTTMDPNNTTQAGIHSAMTHGFSQ	5014	Db	6035	SQPSCHTTPKVGTGSMEDAI VSTSTPGSPETKNVDRDSTSLTPELKEDSTALVMNSTTE	6094
QY	44	-----	43	QY	51	-----	50
Db	5015	LDVTTILMSRIPQDVSWTSPPSVDKTS SPSSFLSSPAMTTPSLISSTLPEDKLSSPMTSLL	5074	Db	6095	SNTVFSSVSLDAATEVSRAEVTVYDPTFMPASAQTKGPDISPESASSHSNPPLTISTH	6154
QY	44	-----	46	QY	51	-----GXYKLDQKLE-----	61
Db	5075	TSGLVKITDILRTRLEPVTSSLNPFSS TSKILATSKDSKDTKEI FPPSINTEETNVKANN	5134	Db	6155	KTIAITQGPVGVTSLGULTDSTIAT SAGTPARTQDVFDSSETTSVMNNDLNDVLKTPS	6214
QY	47	-----	46	QY	62	-----	61
				Db	6215	FSABEANSLSQAPLLVTTSPSVTSTLQEHSTSLSVSVTPVTPLAKITDMDTNLEPV	6274

QY	62	-----	98	-----	97
Db	6275	TRSPQNLRLATSEATTDTHMHPINTAMANGTSSPNBFFYFTVSPDSOPYKATSAV	98	-----	97
QY	62	-----	98	-----	97
Db	6335	VITSTSGDSIVTSMRPSAMKIESETTFSLIPRLRETSTQKIGSSSDTSTVPDKAFT	98	-----	97
QY	62	-----	98	-----	97
Db	6395	AATTEVSRTELTSRSSRTSIQGTETKPTMSPDTSRSTVMTSLFAGLTKSEERTIAQTGPH	98	-----	97
QY	62	-----	98	-----	97
Db	6455	RATSOGLTWDTSITTSQAGTHSAMTHGFSQDLSTLTSRVPYISGTSPPSVEKTS	98	-----	97
QY	62	-----	98	-----	97
Db	6515	SLLSLPAITSPSPVPTTLPESRPSSPVHLTSLPTSGLVKTTDMLASVSLPNLGS	98	-----	97
QY	62	-----	98	-----	97
Db	6575	IPTTSEDIKDTEKMPSTNIATNVGTTTSEKESYSSVPAYSEPCKVTSFPMVTSNIRDT	98	-----	97
QY	62	-----	98	-----	97
Db	6635	IVSTSPGSGSEITRIEMESTFSLAHLGKTSTSQDPIVSTEKSAVLHLKLTTCAT	98	-----	97
QY	67	-----	98	-----	97
Db	6695	VASSRRTSIPGPDHSTESPDISTEIVPSLPISLGITESNNMTIIRTGPPLGST	98	-----	97
QY	67	-----	98	-----	97
Db	6755	LDTPTSSRAGTHSMATQBFPHSEMTVMNKDPEILLWTIPPSIEKTSFSSSLM	98	-----	97
QY	67	-----	98	-----	97
Db	6815	SPVSVSLPKTIHTTPSPMTSLTLPVLMTDTDLGTSPEPTTSSPDLNSTSHVIL	98	-----	97
QY	67	-----	98	-----	97
Db	6875	DTTAEMHPSTSTAAENVETTCGHSQSSVLTDSEKTKATAPMDTTSMTGHTT	98	-----	97
QY	75	NVSKQ	98	-----	97
Db	6935	SVSETTKIKRESTYSLTPLRETSISONASFDTISVLSEVPTGTTAEVSRTEVTSSG	98	-----	97
QY	80	-----	98	-----	97
Db	6995	RTSIPGQSOTVLPLEISTRMTLRLPASPTMTBSAEMTIPTQTGSGSTSQDTLT	98	-----	97
QY	80	-----	98	-----	97
Db	7055	KSQAKTHSTLQRFPHSEMTLMSRPGDMQSSPSLENPSLPLSLPATTSPPPIS	98	-----	97
QY	80	-----	98	-----	97
Db	7115	STLPVTISSPLPVTSLTSSPVTTTMDLHTSPELVTSPPKLSHTSDERLTGKDT	98	-----	97
QY	89	E	98	-----	97
Db	7175	EAVHPSTNTAASNVBIPSGHESPSSALADSETSKATSPMFTSQEDTTVAIST	98	-----	97
QY	90	-----	98	-----	97
Db	7235	TSRIQKESISLSPKLRETGSSVETSSAIEISAVLSEVIGATTEISRTEVTSS	98	-----	97
QY	90	-----	98	-----	97
Db	7295	GSAESTMLPEISTTRKIKRFPSTPILABSEMTIKTQTSPPGSTSESTTLDTST	98	-----	97

QY	98	-----	97	-----	97
Db	7355	ITHSTMTQRLPHSEITTLVSRGAGDVPRPSPPLVEETSPSSQLSLSAMISPPV	98	-----	97
QY	98	-----	97	-----	97
Db	7415	ASSHSSASVTSPLTPGQVKTEVLDAEAPETSPFSSLSSTSVETILATSEVTTD	98	-----	97
QY	98	-----	97	-----	97
Db	7475	PPPNTAVTKVGTSSGHESSPSPVLPDSETTKATSAMGTISIMGDTSVSTLPAL	98	-----	97
QY	98	-----	97	-----	97
Db	7535	QSEPASSLTTLRLRETSTSEETSLATEANTVLSKVGTGATTEVSRTEAIS	98	-----	97
QY	98	-----	97	-----	97
Db	7595	STMSQDISIGTIPRISASSVLTESAKMTITTTQGPSESTLESTLNLTATT	98	-----	97
QY	106	-----	97	-----	97
Db	7655	VIQGFPHPEMTTSMGRGPGVSWPSPFPVKETSPSSPLSLPAVTSHPVSTFLA	98	-----	97
QY	106	-----	97	-----	97
Db	7715	SPLPVTSLTSGPATTTDILGTSTBPTGSSSSSLTTSHERLTYKYDTAHTA	98	-----	97
QY	106	-----	97	-----	97
Db	7775	GGTNVATTSSGYKSSSVLADSSPMCTTSTMGDTSVLTSTPAFLTRRIQTE	98	-----	97
QY	106	-----	97	-----	97
Db	7835	LRSSGSECTSGTKMSTVLSKVPTGATTEISKEDVTSIPGPAQTSISPD	98	-----	97
QY	111	-----	97	-----	97
Db	7895	TSPVMTSBAEITMNTHTSPLGATTQGTSTLDTSTSTSLTMTHTSTISQ	98	-----	97
QY	111	-----	97	-----	97
Db	7955	GPEDVSWSPPLLEKTRFSPSLMSSPATTSPPVSVSTLPESISSPLPVT	98	-----	97
QY	111	-----	97	-----	97
Db	8015	TDMLHKSSEPVNTSPANLSSTSVETLATSEVTTDTEKTHPSSNRRTVD	98	-----	97
QY	111	-----	97	-----	97
Db	8075	FVLADQTSKVTSKVTSPMVTSTMEDTSVSTSTPGFETSRITQETPSLT	98	-----	97
QY	111	-----	97	-----	97
Db	8135	SLATEMTSVLSCVPTGATAEVSRTVETSSRTSISGFAQLTVSPETSTET	98	-----	97
QY	111	-----	97	-----	97
Db	8195	TESAEWMIKTQDPPGSTPESTHTVDISTTPNWEVTHSTVTQRFHSEMT	98	-----	97
QY	111	-----	97	-----	97
Db	8255	LWPSQSVETSSASSLLSLPATTSPSPVSVSTLVEDFPASLPVTSLLT	98	-----	97
QY	111	-----	97	-----	97
Db	8315	ISREPGTSTSNLSTSHERLTTLEDVDTDEAMQSTHTAVTNVRTSISG	98	-----	97
QY	111	-----	97	-----	97
Db	8375	SETPKATSMGTTYTNGETSVSISTDFETSRVQIEPTSSLTSGRETS	98	-----	97
QY	117	-----	97	-----	97

Db 8435 GSTVLSEVPKGATTEVSRTEVISSRGTSMSGPDQFTISPDISTEAIIRLSTPIMTESAE 8494
QY 117 ----- 116
Db 8495 SAITITGSPGATSEGLTLDITTTFMGTHSTASPGFHSBEMTTLMSTRPGDVPWPSL 8554
QY 117 ----- 116
Db 8555 PSVEERASSVSSLSSPAMTSTSPFNSALPESISSPHPVTHLITLGPVKTDMRLTSSBPE 8614
QY 117 ----- 116
Db 8615 TSSPPNLSSTSAEILATSEVTKOREKIHPSSNTPVVVGTVIYKHLSPSSVLADLVTKP 8674
QY 117 ----- 116
Db 8675 TSPMATTSTLGNTSVSTPAPPETMMQTPSTSLTSGLEIRETSQETSATERSASLSGM 8734
QY 117 ----- 116
Db 8735 PGCATTKVSRTEALSIGRTSTPGPAQSTISPELSTETITRISTPLTTGSAEMTTPKTG 8794
QY 117 ----- 116
Db 8795 HSGASSQGTFTLDTSSRASWPGTHSAAHRSPHSGMTPPMSRGPDVSWPSPRSVEKTSP 8854
QY 117 ----- 116
Db 8855 PSSLVLSAVTSPSPLYSTPSSSHSPLRVLTSLFPPVMKTTMDLDTSLBVPVTTSPPSM 8914
QY 117 ----- 122
Db 8915 NITSDSLATSKATMETEALQLSENTAVTQMGTISARQEFYSSYPGLPEPSKVTSPPVTS 8974
QY 123 ----- 122
Db 8975 STIKDIVSTTIPASSEIRIEMESTSTLTPTPRETSTOEIHSATKXPSTVPYKALTSATI 9034
QY 123 ----- 122
Db 9035 EDSMTQVMSSSRGSPDQMSODISTEVITRLSTSPIKAESTEMTITQTGSPGATSRG 9094
QY 123 ----- 122
Db 9095 TLTLDSTTFMSTGTHSTASQFHSQMTALMSRTPGDVPWLHSPSVEEASASFSLSPPV 9154
QY 123 ----- 122
Db 9155 MTSSSPVSTLPDSIHSSSLPVTSLTSLGLVKTTELLGTSSPETSPPNLSSTSAELA 9214
QY 123 ----- 130
Db 9215 TTEVTTDTEKLEMTNVVTSYTHESPSSVLADSVTTKATSSMGITYPTGDTNVLSTPAF 9274
QY 131 ----- 130
Db 9275 SDTSRIQTKSLITPGLMETSISEETSSATEKSTVLSSVPTGATTEVSRTEAISSSRTS 9334
QY 131 ----- 130
Db 9335 IPGPAQSTMSSDTSMETITRISTPLTRKSTDMAITPKTGPGGATSGGTFTLDSSTASW 9394
QY 131 ----- 130
Db 9395 PGTHSATQRPQSVVTPMARGPDVSWPPLSVEKNPSPSSIVSSSVTSPLYSTP 9454
QY 131 ----- 130
Db 9455 SGGSHSPVPVTSFTSIMMKATMDLADLEPETTSAPNMNITSDSLATSKATTETAI 9514
QY 131 ----- 130

Db 9515 HVPENTAASHVETTSATEEELYSPPGFSEPTKVILSPVVTSSIRDNMVSTMPGSSGTR 9574
QY 131 ----- 130
Db 9575 IEIESMSLTPGLRETRTSQDITSTSTSTVLYKMSSGATPEVSRTEVMPSSRSTIPQPA 9634
QY 131 ----- 130
Db 9635 QSTMSLDISDEVVTRLSTSPIMTESAEITTTQTYSLATSQVTLPLGTSMFLSGTHST 9694
QY 131 ----- 130
Db 9695 MSQGLSHSEMNLMSRGPESLSWTSPPREVTRSSSSLSLPLTSLSPVSSSTLLDSSPS 9754
QY 131 ----- 130
Db 9755 SPLPVTSLILPGLVKTTEVLDTSSEPKTSSPNLSSTSVETPATSEIMTDEKIHPSNT 9814
QY 131 ----- 130
Db 9815 AVAKVRTSSSVHSHSVLADSETTITIPSMGITSAVDDTTVTNPAFSETRRIPEPT 9874
QY 131 ----- 130
Db 9875 FSLTPGPRETSTSEBTTSITETSAVLYGVPTSATTEVSMTEIMSNRTHIPDSQSTMS 9934
QY 131 ----- 140
Db 9935 DIITEVITRLSSSMSESTQMTITTKQSPGATAQSTLTILATTAPLARHTSTVPPRFL 9994
QY 141 ----- 140
Db 9995 HSEMTLMRSRPENPSWKSPFVEKTSSSSLLSLPVTSPSVSTLPQSIPSSSFVTS 10054
QY 141 ----- 140
Db 10055 LLTPGMVKTDTSTPEGTSLSPNLSGTSVEILAASEVTTDEKIHPSSSMAVTVNGTSS 10114
QY 141 ----- 143
Db 10115 GHELYSSVSIHSEPSKATYPVGTGTPSSMAETISITSMANPETTGPEAFPFHLSGTGFRKT 10174
QY 144 ----- 143
Db 10175 NMSLDTSVPTNTNTPSPGSGTHLQSSKTDFTSSAKTSSPDWPPASQYTEIPVDIITPN 10234
QY 144 ----- 143
Db 10235 ASPSITESTGITSPPESRFTMSVSTHLSLTDLLPSAETISTGTVMPSLSEAMTSFAT 10294
QY 144 ----- 143
Db 10295 GVPRAISGSGSPFRSTESPGDATLSTIAESLPSTPVPFSSSTFTTTSDSTIPALHEIT 10354
QY 144 ----- 143
Db 10355 SSSATPYRVDTSLGTESSSTTEGLVNVSTLDTSSQPGRTSSPTILDTRMTESVELGTVTS 10414
QY 144 --RVPSM----- 148
Db 10415 AYQVPSLSTRLTRDGIIMEHITKI PNEAAHRTIRPVKQPOTSTGPASPKLHGTGTRM 10474
QY 149 -----TCLEML----- 155
Db 10475 ETTTALKTTTTALKTTGRATLTTVVYPTLGLTPLNASROMASTILTEMITPPYVPP 10534
QY 156 ----- 155
Db 10535 DVPETTSSLATSLGAETSTALPRTPSVLNRSETTASLVSRSGAERSPVIQTLDVSSSE 10594
QY 156 -----NVSK----- 159
Db 10595 POTTASWVIHPAETIPTVSKTTPNFHSELDTVSSSTATSHGADVSSAIPNTNISPELDAL 10654


```
QY 160 ----- 159
Db 10655 TPLVTSIGTDSTTFPLTKSPHETETRTTTLWTHPAETSSITPRTIPNFSHIESDATPSI 10714
QY 160 ----- 161
Db 10715 ATSPCAETSSAIPIMTVSPGAEDLVTSQVTSSTGDRNWTPTLTLSPGEPKTIASLVTHP 10774
QY 162 ----- 161
Db 10775 EAQTSSAIPSTISPAVSLVTSVMVTSIAAKTSTTNEALTNSPCEPATTVSLVTHPAQTS 10834
QY 162 ----- 161
Db 10835 PTVPMWTTIFPHSKSDTTPMTTSHGAESSAVPTPTVTEVPGVVPLVTSSRAVISTT 10894
QY 162 ----- 161
Db 10895 IPILTLSPGEPETTPSMATSHGTEAGSAIPTPTVSPGVGVVTSVTSRAVTSITPIL 10954
QY 162 ----- 161
Db 10955 TFSLGEPEETTPSMATSHGTEAGSAIPTVLPVPCWMTSLVASSRAVTSITLTLTSPGE 11014
QY 162 ----- 161
Db 11015 PETTPSMATSHGAESSTVPTVSPVGVVTSVTSVSSGVNSTIPTLILSPGELETTPS 11074
QY 162 ----- 161
Db 11075 MATSHGAESSAVPTPTVSPGVGVVPLVTSRAVTSITPILTLSSSEPETTPSMATS 11134
QY 162 ----- 161
Db 11135 HGVEASSAVLTVSPVPCWMTSLVTSRAVTSITPILTLISSDEPETTSLVTHSEAKMI 11194
QY 162 ----- 161
Db 11195 SAIPTLAVSPVQGLVTSVTSVSSGSETSAFSLNLTVASSQPETIDSWVAHPGTEASSVPT 11254
QY 162 ---DTGE----- 165
Db 11255 LTVSTGEPTFNI SLVTHPAESSLPRTRTSRFGHSELDTWMPSTVTSPEAESSAISTTIS 11314
QY 166 ----- 165
Db 11315 PGIPGVLTSLVTSSSGRDISATFTVPVSPESPHESATASWTHPAVTSITVPTPTPNYSHSE 11374
QY 166 ----- 165
Db 11375 PDTPPGIATSPGAEATSDPPTITVSPDVPDMVTSQVTSSTGDTFSITPILTLSSGPEIT 11434
QY 166 ----- 174
Db 11435 TSPITYETHSSAIPLVPSPGASKMLTSLVSSGDTSTTPTTLETPEYBETTAIOL 11494
QY 175 ----- 178
Db 11495 IHPAETNTWVPTTPKFSHKSDTTLPAVITSPGPEASSAVSTTISPDMSDLVTSLVPS 11554
QY 179 ----- 178
Db 11555 SGTDTSTTPTLSETPEPETTATLWTHPAETSTTVSGTIPNFSHRGSDTAPSMVTSPOV 11614
QY 179 ----- 179
Db 11615 DTRSGVPTTIPPSIPGVVTSQVTSATDTSTAIPLTLPSPGEPETTASSATHPGTQGF 11674
QY 180 NVP----- 182
Db 11675 TVPIRTVPSSEPDTMASWTHPPQTPTVSRRTSSPSHSSPDATPVNATSPREASSAVL 11734
QY 183 ----- 182
Db 11735 TTISPGAPENVTSQITSSGAATSTTVPTLTHSPGPETTTALLSHTRTETSKTFPASTVF 11794
QY 183 ----- 182
Db 11795 PQVSETTASLTIRPGAETSTALPTQTTSSLFTLLVTGTSRVDLSPTASPGVSAKTAPLST 11854
QY 183 ----- 182
Db 11855 HPGTETSTMIPTSTLSIGLLETTGLLATSSAETSTSTLTLTVSPAVSGLSSASITTDKP 11914
QY 183 ---RWTK----- 187
Db 11915 QVTSMNTETSPSVTSVGPPEFSRTVTGTTMLIPSEMPTPKTSHGEGVSPVTILRTTM 11974
QY 188 ----- 187
Db 11975 VEATNLATGSSPTVAKTTTFTNTLAGSLFTPLTTPGMSTLASESVTSRTSYNHRSWIST 12034
QY 188 ----- 187
Db 12035 TSSYNRRYWPATSTPVTSTFSPGISTSSIPSSSTAATVPFWPPTLNTITINLOYBEDMR 12094
QY 188 ---LYVG-----P 192
Db 12095 HPGSRKFNATERLOGLLKLPLFRNSLEVLYSGRLASLRPEKSSAMAVDAICTHRPDP 12154
QY 193 TKVNDVDSQTIYP----- 204
Db 12155 EDGLDRERLYWELSNLTNGIOELGPYTLDRNSLYNGFTHRSSMPTTTPGTSTVDVGT 12214
QY 205 ----- 204
Db 12215 SCTPSSSPPTAAGPLLMPTLNTITINLOYBEDMRTGSRKENTMESVLQGLLKLPLFKN 12274
QY 205 ----- 204
Db 12275 TSVGPLYSGRLTLRLRPEKGAATGVDAICTHRLDPKSPGLNREQLYWELSKLNDIEEL 12334
QY 205 ----- 204
Db 12335 GPYTLDRNSLYNGFTHQSSVSTTTPGTSTVDLRTSGTSPSSSPTIMAAGPLLVPTL 12394
QY 205 ----- 204
Db 12395 NFTITNLOYGEDMHPGSRKFNTERVLOGLLGPFNKTSVGPLYSGRLTSLSRSEKGA 12454
QY 205 ----- 204
Db 12455 ATGVDAICTHHLDPKSPGLNRRLYWELSQLTNGIKELGPYTLDRNSLYNGFTHRTSVP 12514
QY 205 ---LGLTALLRYAQ----- 216
Db 12515 TSSTPGTSTVDLGTSGTSPFLSPATAGPLLVLTFTLNTITNLYKEDMHGSRKFNNT 12574
QY 217 ----- 216
Db 12575 ERVLQTLGPMFNKTSVGLLYSGRLTLRSEKGAATGVDAICTHRLDPKSPGLDREQL 12634
QY 217 ----- 216
Db 12635 YWELSQLTNGIKELGPYTLDRNSLYNGFTHWIPVPTSSTPGTSTVDLGSSTPSSLPST 12694
QY 217 ----- 216
Db 12695 AAGPLLVPTLNTITINLOYBEDMHPGSRKFNTERVLOGLLGPFNKTSVGLLYSGCR 12754
QY 217 ---RN--- 218
Db 12755 LTLRSEKGAATGVDAICTHRLDPKSPGVDRQLYWELSQLTNGIKELGPYTLDRNSLY 12814
QY 219 ----- 218
```

Db	12815	VNGFTHQTSAPNTSTPGTSTVDLGTSGTSPSLPSTAGPLLVPTFLNFTITNLQYEEDM	12874	Db	13895	LGPYXLDKRXSLYVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXXTSAGPLLVPTFLN	13954
QY	219	-----CTH---	221	QY	253	-----	252
Db	12875	RHPGRKFNTTERRVLQGLLKPLFKSTSVGPLYSGCRLTLRLRSEKGAATGVDAICTHRLD	12934	Db	13955	FTITNLQYEEDMHHPGSRKFNTTERRVLQGLLGPMPKNTSVGLLYSGCRLTLRLRPEKGAA	14014
QY	222	-----	221	QY	253	-----	252
Db	12935	PKSPGVDRQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHQTSAPNTSTPGTSTVDLG	12994	Db	14015	TGMDAICSHRLDPKSPGLDRQLYWELSQLTHGIKELGPYTLDRNSLYVNGFTHRSSVAP	14074
QY	222	-----SFVLVN-----	227	QY	253	-----KPK-----	255
Db	12995	TSCTPSSLPSPSTAGPLLVPTFLNFTITNLQYEEDMHHPGSRKFNTTERRVLQGLLGPMPK	13054	Db	14075	TSTPGTSTVDLGTSGTSPSSLPSPPTAVPLLVPTFLNFTITNLQYEDMRHFGSRKFNTTE	14134
QY	228	-----	227	QY	256	-----QAP-----	258
Db	13055	NTSVGLLYSGCRLTLRLRPEKNGAATGMDAICSHRLDPKSPGLNRQLYWELSQLTHGIKE	13114	Db	14135	RVLQGLLGPLFKNSVGPLYSGCRLISLRSEKGAATGVDAICTHHLNPPQSGLDREQLY	14194
QY	228	-----AMSRN-----	232	QY	259	-----VKE-----	261
Db	13115	LGPYTLDRNSLYVNGFTHRSSVAPTSTPGTSTVDLGTSGTSPSLPSPPTAVPLLVPTFLN	13174	Db	14195	WQLSOMTNGIKELGPYTLDRNSLYVNGFTHRSSGLTSTPWTSTVDLGTSPSPSPPT	14254
QY	233	-----LFR-----	235	QY	262	-----	261
Db	13175	FTITNLQYEDMRHHPGSRKFNTTERRVLQGLLGPLFKNSVGPLYSGCRLISLRSEKGA	13234	Db	14255	TAGPLLVPTFLNFTITNLQYEEDMRHHPGSRKFENATERVLQGLLSPIFKNSSVGPLYSGCR	14314
QY	236	-----VPK-----	235	QY	262	-----	261
Db	13235	TGVDAICTHHLNPPQSGLDREQLYWELSQLTNGIKELGPYTLDRNSLYVNGFTHRSSGLT	13294	Db	14315	LTSLRPEKGAATGMDAVCLYHPNPKRPGLDRQLYWELSQLTHNITELGPVSLDRDLSL	14374
QY	236	-----	238	QY	262	-----QPEKKA	267
Db	13295	TSTPWTSTVDLGTSGTSPSPPTAGPLLVPTFLNFTITNLQYEEDMRHHPGSRKFENATE	13354	Db	14375	VNGFTHQSSMTTRTPDSTWHLATSRTPASLSGTTASPLLVLTINCTITNLQYEEDM	14434
QY	239	-----	238	QY	268	KKTQS-----	272
Db	13355	RVLQGLLSPIFKNSSVGPLYSGCRLTSLRPEKGAATGMDAVCLYHPNPKRPGLDRQLY	13414	Db	14435	RTGSRKFNTMESVLQGLLKPLFKNTSVGPLYSGCRLTLRLPKKGAATGVDAICTHRLD	14494
QY	239	-----	238	QY	273	-----	272
Db	13415	WELSQLTHNITELGPVSLDRDLSLYVNGFTHQNSVPTSTPGTSTVYVWATGTPSPPGHT	13474	Db	14495	PKSPGLNRQLYWELSQLTNDIEELGPYTLDRNSLYVNGFTHQSSVSTTSTPGTSTVDLR	14554
QY	239	-----	238	QY	273	-----	272
Db	13475	EPGILLIPTFNFNTITNLHYEENMQHPGSRKFNTTERRVLQGLLKPLFKNTSVGPLYSGCR	13534	Db	14555	TSGTPSSLSPTIMXXXPLLPFTFLNFTITNLXYEEXXPGSRKFNTTERRVLQGLLRPL	14614
QY	239	-----Y	239	QY	273	-----	272
Db	13535	LTSLRPEKGAATGMDAVCLYHPNPKRPGLDRQLYCELSQLTHNITELGPVSLDRDLSL	13594	Db	14615	FKNTSVSSLYSGCRLTLRLRPEKGAATRVDAACTVYRDPDPKSPGLDRQLYWELSQLTHSI	14674
QY	240	ING-----	242	QY	273	-----	272
Db	13595	VNGFTHQNSVPTSTPGTSTVYVWATGTPSPSPPGHTPGPLLPFTFNFNTITNLHYEENM	13654	Db	14675	TELGPYTLDRVSLYVNGFNPRSSVPTSTPGTSTVHLATSGTPSSLPGHTXXXPLLPPT	14734
QY	243	-----	242	QY	273	-----	272
Db	13655	QHPGSRKFNTTERRVLQGLLKPLFKNTSVGPLYSGCRLTLRLRPEKHEAATGVDTICTHRVD	13714	Db	14735	LNFTITNLXYEEXXPGSRKFNTTERRVLQGLLKPLFRNSSLEYLYSGCRLASLRPEKDS	14794
QY	243	-----TKLQNTMRKL-----	252	QY	273	-----	272
Db	13715	PIGGLDRERLYWELSQLTNSITELGPYTLDRDLSLYVNGFNPRSSVPTSTPGTSTVHLA	13774	Db	14795	SAMAVDAICTHRPDPEDLGLDRERLYWELSQLTNGIQELGPYTLDRNSLYVNGFTHRSSF	14854
QY	253	-----	252	QY	273	-----TTTTPYFS-----	279
Db	13775	TSGTPSSLPGHTAPVPLIPFTFLNFTITNLHYEENMQHPGSRKFNTTERRVLQGLLKPLFK	13834	Db	14855	LTTSTPWTSTVDLGTSGTSPSPPTAGPLLVPTFLNFTITNLQYEEDMRHHPGSRRTFT	14914
QY	253	-----	252	QY	280	-----	279
Db	13835	NTSVGPLYSGCRLTLRLRPEKHEAATGVDTICTHRVDPIGGLDRERLYWELSQLTXXIXE	13894	Db	14915	TERVLQGLLTPLFKNTSVGPLYSGCRLTLRLRPEKQBAATGVDTICTHRVDPGPGGLDRR	14974
QY	253	-----	252	QY	280	-----	279
Db	13895	LYWELSQLTNSITELGPYTLDRDLSLYVNGFNWSSVPTSTPGTSTVHLATSGTPSSLP	15034	Db	14975	LYWELSQLTNSITELGPYTLDRDLSLYVNGFNWSSVPTSTPGTSTVHLATSGTPSSLP	15034

```
QY 280 ----- 279
Db 15035 HTAPVLLIPFTLNFTITDLHYEENMHQPGSRKFNTTTERVLOGLLPLFKSTSVGLYSY 15094
QY 280 -----YT----- 281
Db 15095 CRLTLRPEKHGAATGVDAICTLRDPTGGLDRERYWELSQLTNSVTELGYTLDRDS 15154
QY 282 -----TSA-----ALNVT--TNVTY-- 294
Db 15155 LYVNGFTHRSSVPTTSPGTSVAVHLETSPTASLPGHATAGPLLVPTLNFTITNLQYEE 15214
QY 295 ----- 294
Db 15215 DMRHPSGRKFSTTTERVLOGLLPLFKNTSVSSLYSGCRLTLRPEKGAATRVDAVCTHR 15274
QY 295 -----SITTAARVSTSI- 308
Db 15275 PDPKSPGLDRERYWKLSQLTHGIGELGYTLDRHSLYVNGFTHQSSMTTTRTPTSTMH 15334
QY 309 ----- 308
Db 15335 LATSPTASLPGTTASPLLVLTINFTITNQRYEENMHQPGSRKFNTTTERVLOGLLRPV 15394
QY 309 -----AYRPDS- 315
Db 15395 FKNTSVGLYSGCRLTLRPPKDKGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHSI 15454
QY 316 ----- 315
Db 15455 TELGPYTODRDSLYVNGFTHRSSVPTTSPGTSVAVHLETSPTASLPGHATAGPLVPT 15514
QY 316 ----- 315
Db 15515 LNFTITNLQYEEEDMRHPSGRKFNTTTERVLOGLLPLFKSTSVGLYSGCRLTLRPEKRG 15574
QY 316 ----- 315
Db 15575 AATGVDTICTHRLDPLNPLDREQLYWELSKLTRGIIELGPVLLDRGSLYVNGFTHRTSV 15634
QY 316 ----- 315
Db 15635 PTTSTPGTSTVDLGTSPFLSPAPAXXPXLLXPTLNFTITNLXYEEXXPGSRKFNT 15694
QY 316 ----- 315
Db 15695 TERVLOTLGPMFKNTSVGLLYSGCRLTLRSEKGAATGVDAICTHRLDPKSPGVDRQ 15754
QY 316 ----- 315
Db 15755 LYWELSQLTNGIKELGPVTLDRNSLYVNGFTHWIPVPTSSPTGTSVTLGSGTSPSP 15814
QY 316 -----FMKSMATQJR----- 326
Db 15815 TTAGPLLVPTLNFTITNLKYBEDMHCPGSRKFNTTTERVLOGLLPMFKNTSVGLYSGC 15874
QY 327 ----- 326
Db 15875 RLTLRSEKGAATGVDAICTHRLDPKSPGVDRQYWELSQLTNGIKELGPVTLDRNSL 15934
QY 327 -----DLAT----- 330
Db 15935 YVNGFTHQTSAPNTSPGTSVTLGTSPTSPSSLPSTXXXPLXPTLNFTITNLXYEEX 15994
QY 331 ----- 330
Db 15995 MXXPGSRKFNTTTERVLOGLLXPKFKXTSVGXLYSGCRLTLRXXKXAAATXVDXXCXXX 16054
QY 331 -----WV----- 332
Db 16055 DPXXPGLDREXLYWELSKLTXIXELGPYXLDXSLYVNGFTHWIPVPTSPGTSVTL 16114
QY 333 -----YTLRYRONPFCBPS----- 347
Db 16115 GSGTSSLSPTTAGPLLVPTLNFTITNLKYBEDMHCPGSRKFNTTTERVLOGLLPMFK 16174
QY 348 ----- 347
Db 16175 NTSVGLYSGCRLTLRSEKGAATGVDAICTHRLDPKSPGVDRQYWELSQLTNGIKE 16234
QY 348 -----RNRTAVSEF----- 356
Db 16235 LGPYTLDRNSLYVNGFTHQTSAPNTSPGTSVTLGTSPTSPSSLPSTAGPLLVPTLN 16294
QY 357 -----MKNTHV----- 362
Db 16295 FTITNLQYEEEDMHQPGSRKFNTTTERVLOGLLPMFKNTSVGLLYSGCRLTLRPEKNGAA 16354
QY 363 ----- 362
Db 16355 TGMDAICTHRLDPKSPGLDREXLYWELSKLTXIXELGPYXLDXSLYVNGFXXXXXXX 16414
QY 363 ----- 362
Db 16415 TSTPGTSVXLXTSGTPXXXPKXXXPLXPTLNFTITNLKYEEXXPGSRKFNTTE 16474
QY 363 -----LIRNET----- 368
Db 16475 RVLOGLLPLFRNSSLEYLYSGCRLASLRPEKSSAMAVDAICTHROPDGLDRERLY 16534
QY 369 -----PVTI----- 372
Db 16535 WELSNLTNGIQELGPVTLDRNSLYVNGFTHRSMPTTSTPGTSVTVGTSPTSPSSPSPT 16594
QY 373 -----YG----- 374
Db 16595 TAGPLLIPTLNFTITNLQYGEDMHGHPGSRKFNTTTERVLOGLLGPFKNTSVGLYSGCR 16654
QY 375 -----TLDMSLY 382
Db 16655 LTSRSEKGAATGVDAICIIHLDPKSPGLNRERLYWELSQLTNGIKELGPVTLDRNSLY 16714
QY 383 YN----- 384
Db 16715 VNGFTHRTSVPTTSTPGTSVTLGTSPTSPATAGPLLVLTINFTITNLKYEEDM 16774
QY 385 ----- 384
Db 16775 HRPGSRKFNTTTERVLOTLGPMFKNTSVGLLYSGCRLTLRSEKGAATGVDAICTHRLD 16834
QY 385 ----- 384
Db 16835 PKSPGLDREXLYWELSKLTXIXELGPYXLDXSLYVNGFXXXXXXXSTPGTSXVXLX 16894
QY 385 ----- 384
Db 16895 TSGTPXXXPKXTXXXPLXPTLNFTITNLXYEEXXPGSRKFNTTTERVLOGLLRPVK 16954
QY 385 ----- 384
Db 16955 NTSVGLYSGCRLTLRPPKDKGAATKVDAICTYRPDPKSPGLDREQLYWELSQLTHSITE 17014
QY 385 -----ETMFVENKTASDSNKT----- 401
Db 17015 LGPYTODRDSLYVNGFTHRSSVPTTSPGTSVAVHLETTGTSFPGHTBPGPLIPTFN 17074
QY 402 ----- 401
Db 17075 FTITNLRYEENMHQPGSRKFNTTTERVLOGLLPLFKNTSVGLYSGCRLTLRPEKQAAA 17134
QY 402 ----- 401
Db 17135 TGVDTICTHRLDPIGGLDRERLYWELSQLTNSITELGPYTLDRDSLYVDGPNWSSVPT 17194
QY 402 ----- 401
```

Db	17195	TSTPGTSTVHLATSGTSPPLPGHTAPVPLLIPTFLNFTITDLHYEENMOHPSGRKFNTTE	17254	Db	18275	VNGFTHQNSVPTTSTPGTSTVYWAATTGTPSSPFGHTPEGPLLIPTFTNFTITNLHYEENM	18334
QY	402	-----PTSPSMG-----	408	QY	432	-----	431
Db	17255	RVLQGLLKPLFKSTSVGPLYSGCRLTLRLRPEKHGAATGVDAICTLRDPTGPGDRERLY	17314	Db	18335	QHPGSRKFNTTERTVLOGLLTPLFKNTSVGPLYSGCRLTLRLRPEKQEAATGVDITCTHRVD	18394
QY	409	-----	408	QY	432	-----	431
Db	17315	WELSQLTNSITELGPYTLDRDSLTVNGFNPWSSVPTTSTPGTSTVHLATSGTPSSLPGHT	17374	Db	18395	PIGPGLDREXLYWELSXLTXIXELGPYXLDRLXSLVNGFXXXXXXXTSTPGTSXVXLX	18454
QY	409	-----FORTFIDPL-----	417	QY	432	-----	431
Db	17375	TAGPLLVPTFLNFTITNLKYEEDMCPGSRKFNTTERTVLOQLHGPMPKNTSVGPLYSGCR	17434	Db	18455	TSGTPXXXPXTXXXPLLPFTLNFTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPFK	18514
QY	418	-----WD-----	420	QY	432	-----	431
Db	17435	LTLRSEKOGAATGVDAICTHRLDPKSPGLDREXLYWELSXLTXIXELGPYXLDRLXSLY	17494	Db	18515	XTSVGXLVSGCRLTLRLRPEKXXAATVVDXXCXXDXDPXYPGLDREXLYWELSXLTXIXE	18574
QY	421	-----	421	QY	432	-----	431
Db	17495	VNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXXPLLPFTLNFTITNLXYEEXM	17554	Db	18575	LGPYXLDRLXSLVNGFTHRESSVPTTSSPGTSTVHLATSGTPSSLPGHTAPVPLLIPTFLN	18634
QY	422	-----	421	QY	432	-----	431
Db	17555	XXPGSRKFNTTERTVLOGLLXPFKXTSVGXLVSGCRLTLRLRPEKXXAATVVDXXCXXDX	17614	Db	18635	FTITNLHYEENMOHPSGRKFNTTERTVLOGLLKLPLFKSTSVGPLYSGCRLTLRLRPEKHGA	18694
QY	422	-----DSL-----	424	QY	432	-----	431
Db	17615	PXXPGLDREXLYWELSXLNSITELGPYTLDRDSLTVNGFTHRSSMPTTSTPGTSVAHLE	17674	Db	18695	TGVDAICTLRDPTGPGDLREXLYWELSXLTXIXELGPYXLDRLXSLVNGFXXXXXXXT	18754
QY	425	-----LF- 426	426	QY	432	-----	431
Db	17675	TSGTPASLPCHTAPGPLVPFTLNFTITNLQVEEDMRHPSGRKFNTTERTVLOGLLKPLFK	17734	Db	18755	TSTPGTSXVXLXTSGTPXXXPXTXXXPLLPFTLNFTITNLXYEEXMXXPGSRKFNTTE	18814
QY	427	-----LDEIR-----	431	QY	432	-----	431
Db	17735	STSVGPLYSGCRLTLRLRPEKGAATGVDITCTHRLDPLNPGDLREXLYWELSXLTXIXE	17794	Db	18815	RVLQGLLXPFKXTSVGXLVSGCRLTLRLRPEKXXAATVVDXXCXXDXDPXYPGLDREXLY	18874
QY	432	-----	431	QY	432	-----	431
Db	17795	LGPYXLDRLXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXXPLLPFTLN	17854	Db	18875	WELSXLTXIXELGPYXLDRLXSLVNGFTHRTSTVPTTSTPGTSTVHLATSGTPSSLPGHT	18934
QY	432	-----	431	QY	432	-----NF-----	433
Db	17855	FTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPFKXTSVGXLVSGCRLTLRLRPEKXXA	17914	Db	18935	APVPLLIPTFLNFTITNLQVEEDMRHPSGRKFNTTERTVLOGLLSPIFKNSSVGPLYSGCR	18994
QY	432	-----	431	QY	434	-----SLR-----	436
Db	17915	TXVDXXXXXXDPGPGDLREXLYWELSXLTXIXELGPYXLDRLXSLVNGFHPRSSVPT	17974	Db	18995	LTSLRPEKOGAATGMDAVCLYHPNPKRPGLDREQLYCELSQLTHNITELGPYSLDRDSL	19054
QY	432	-----	431	QY	437	-----	436
Db	17975	TSTPGTSTVHLATSGTPSSLPGHTAPVPLLIPTFLNFTITNLHYEENMOHPSGRKFNTTE	18034	Db	19055	VNGFTHQNSVPTTSTPGTSTVYWAATTGTPSSPFGHTXXXPLLPFTLNFTITNLXYEEXM	19114
QY	432	-----	431	QY	437	-----	436
Db	18035	RVLQGLGPMFKNTSVGLLYSGCRLTLRLRPEKNGAATGMDAICSHRLDPKSPGLDREXLY	18094	Db	19115	XXPGSRKFNTTERTVLOGLLXPFKXTSVGXLVSGCRLTLRLRPEKXXAATVVDXXCXXDX	19174
QY	432	-----	431	QY	437	-----	436
Db	18095	WELSXLTXIXELGPYXLDRLXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXT	18154	Db	19175	PXXPGLDREXLYWELSXLTXIXELGPYXLDRLXSLVNGFTHWSGLTSTPTWTSTVDLG	19234
QY	432	-----	431	QY	437	-----SPT-----	439
Db	18155	XXXPLLPFTLNFTITNLXYEEXMXXPGSRKFNTTERTVLOGLLXPFKXTSVGXLVSGCR	18214	Db	19235	TSGTPSPVPTTAGPLVPFTLNFTITNLQVEEDMRHPSGRKFNAATERVLOGLLSPIFK	19294
QY	432	-----	431	QY	440	-----	439
Db	18215	LTLRPEKXXAATVVDXXCXXDXDPXYPGLDREXLYWELSXLTXIXELGPYXLDRLXSLY	18274	Db	19295	NTSVGPLYSGCRLTLRLRPEKQEAATGVDITCTHRVDPGPGDLREXLYWELSXLTXIXE	19354
QY	432	-----	431	QY	440	-----	439
				Db	19355	LGPYXLDRLXSLVNGFXXXXXXXTSTPGTSXVXLXTSGTPXXXPXTXXXPLLPFTLN	19414

QY	440	-----	439	QY	446	-----	445
Db	19415	FTITNLXYEEXXPGSRKPHRRNTTTERVLOGLLXPFXKTSVGLYSGCRLTLRXXEKX	19474	Db	20495	HTAPGELLVPFTLNTITNLQYEDMRHPSGRKFNTERVLOGLLKPFPKSTSVGPLYSG	20554
QY	440	YN-----	442	QY	446	-----DEHRA-----	451
Db	19475	XAATXVDXXCXXXPGLDREXLYWELSLTXIXLGPYXLDXSLYVNGFTHRSF	19534	Db	20555	CRLLTLRPEKGAAGVDTICTHRLDPLNPGLDREQLYWELSKLTGIIELGPYLLDRGS	20614
QY	443	-----	442	QY	452	-----	451
Db	19535	GLTTSPTWSTVDLGTSGTSPVPSPPTAGPLLVPTLNTITNLQYEDMRHPSGRKEN	19594	Db	20615	LYVNGFTHRNFPITSTPGTSTVHLGTSETSSLPRIVPGLLVPTLNTITNLQYEE	20674
QY	443	-----LTP-----	445	QY	452	-----	451
Db	19595	TERVLOGLLTLFRNTSVSLYSGCRLTLRPEKGAATRVDAVCTHRPDPKSPGLDRE	19654	Db	20675	AMRHPSGRKFNTERVLOGLLRPLFKNTSIGPLTYSSCRLTLRPEKKAATRVDAICT	20734
QY	446	-----	445	QY	452	-----	451
Db	19655	XLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXXXTSTPGTSKXVLTSGTXXXXP	19714	Db	20735	HHDPQSPGLNREQLYWELSQLTHGITELGPTLDRDSLVDGFTHWSPIPTSTPGTGI	20794
QY	446	-----	445	QY	452	VNLSTS-----	457
Db	19715	XXTXXPPLXPFTLNTITNLXYEEXXPGSRKFNTERVLOGLLXPFXKTSVGLXYS	19774	Db	20795	VNLGTSGIPPSLPETTTXXXPLLPFTLNTITNLXYEEXXPGSRKFNTERVLOGLLK	20854
QY	446	-----	445	QY	458	-----	457
Db	19775	GCRLTLRXXEKKAAATXVDXXCXXXPGLDREXLYWELSLTXIXELGPYXLDX	19834	Db	20855	PLFKSTSVGPLYSGCRLTLRPEKGVATRVDAICTHRPDKPIPGLDROQLYWELSQLTH	20914
QY	446	-----	445	QY	458	-----	457
Db	19835	SLYVNGFTHWIPVPTSTPGTSTVDLGTSGTSPSPPTAGPLLVPTLNTITNLQYGE	19894	Db	20915	SITELGPYTLDRDSLTVNGFTORSSVPTTSTPGTFTVQPETSETPSSLPCTATGVLIP	20974
QY	446	-----	445	QY	458	-----	457
Db	19895	DMHPSGRKFNTERVLOGLLGPIFKNTSVGLYSGCRLTLRSEKDGAAATGVDAICHH	19954	Db	20975	FTLNTITNLQYEDMRHPSGRKFNTERVLOGLLMPFKNTSVSSLYSGCRLTLRPEK	21034
QY	446	-----	445	QY	458	-----	457
Db	19955	LDPKSPGLDREXLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXXXTSTPGTSXVK	20014	Db	21035	DGAATRVDAVCTHRPDKSPGLDRERLYWELSQLTHGITELGPTLDRHSLYVNGFTHOS	21094
QY	446	-----	445	QY	458	-----	457
Db	20015	LXTSGTXXXXXXTXPLLPFTLNTITNLXYEEXXPGSRKFNTERVLOGLLXP	20074	Db	21095	SMTTRTPTDSTMHLATRTASLPGTPTASPLLVLTINFTINTRYENMHHPGSRKP	21154
QY	446	-----	445	QY	458	-----	457
Db	20075	FKXTSVGLYSGCRLTLRXXEKKAAATXVDXXCXXXPGLDREXLYWELSLTXIX	20134	Db	21155	NTTERVLOGLLRPVFKNTSVGPLYSGCRLTLRPPKOGAATKVAICTYRPPDKSPGLDR	21214
QY	446	-----	445	QY	458	-----	457
Db	20135	XELGPYXLDXSLYVNGFTHQIFAPNTSTPGTSTVDLGTSGTSPSPSPTAGPLLVPT	20194	Db	21215	EQLYWELSQLTHSITELGPTQDRDSLTVNGFTORSSVPTTSPVGTPTVDLGTSGTPVSK	21274
QY	446	-----	445	QY	458	-----	457
Db	20195	LNFTITNLQYEDMRHPSGRKFNTERVLOGLLGPMPFKNTSVGLLYSGCRLTLRPEKNG	20254	Db	21275	PGPSAASPLLVLTNGTITNLRYEENMHHPGSRKFNTERVLOGLLRSLFKSTSVGPLY	21334
QY	446	-----	445	QY	458	-----	457
Db	20255	AATRVDAVCTHRPDKSPGLDREXLYWELSLTXIXELGPYXLDXSLYVNGFXXXXXX	20314	Db	21335	SGCRLTLRPEKDGATGVDAICTHHPDKSPRLDREQLYWELSQLTHNITELGHVALDN	21394
QY	446	-----	445	QY	458	-----	457
Db	20315	XXTSTPGTSKXVLTSGTXXXXPXTAPVPLLPFTLNTITNLHYEENMHHPGSRKFN	20374	Db	21395	DSLTVNGFTHRSSVPTTSTPGTPTVVLGASKTPASIFGPSAASHLLILFTLNTITNLRY	21454
QY	446	-----	445	QY	458	-----	457
Db	20375	TERVLOGLLRPLFKSTSVGPLYSGCRLTLRPEKGAATGVDAICTLRDPTGPGLDRE	20434	Db	21455	EENMHPSGRKFNTERVLOGLLRPLFKNTSVGPLYSGSRLTLRPEKDGATGVDAICTH	21514
QY	446	-----	445	QY	458	-----	457
Db	20435	LYWELSQLTNSVTELGPYTLDRDSLTVNGFTORSSVPTTSTPGTSAVHLETSPTASLPG	20494	Db	21515	RPDPTGPGLDREQLYLELSQLTHSITELGPTLDRDSLTVNGFTHRSSVPTTSTGVVSEB	21574
				QY	458	-----	457

Db 21575 PFTLNTINNLRYMADMGQSGSLKFNITDNNVMKHLSPLFQSSIGARYTGCRVIALRSV 21634
QY 458 -----
Db 21635 KNGAETRVDLLCTYLQPLSGPLPIKQVPHLSQQTHGTRGLGPKSLDKDSLNGYNEP 21694
QY 458 -----
Db 21695 GLDEPPTPKPATFTPLPLSEATTAMGYHLKTLTLNFTISNLIQYSPDMGKSATFNSTEG 21754
QY 458 -----NSLW 462
Db 21755 VLQHLRLPLFKQSMGPFYLGQLISLRPEKDAATGVDTTCTYHPDPVPGGLDIOQLYW 21814
QY 463 -----WL 464
Db 21935 TLNASFHWL 21943

RESULT 10
AAO29660
ID AAO29660 standard; protein; 8601 AA.
AC AAO29660;
DT 03-SEP-2003 (first entry)
DE Paederus fuscipes Pedf mixed type I PKS/NRPS.
KW Pederin; polyketide synthase; PKS; nonribosomal peptide synthase; NRPS;
KW antitumour agent; beetle; Pedf.
OS Paederus fuscipes.

Key Location/Qualifiers
Misc-difference 7136 /label= Unknown
/note= "Encoded by RTC"
WO2003044186-A2.
30-MAY-2003.
21-NOV-2002; 2002WO-EP013085.
22-NOV-2001; 2001EP-00127395.
(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
Piel J;
WPI; 2003-468769/44.
N-PSDB; AAL60188.
Novel gene cluster encoding modular polyketide synthase enzyme involved
in biosynthesis of antitumor compound pederin, useful in preparing
modified pederin biosynthesis gene cluster, or a modified pederin
molecule.
Claim 10; Page 46-69; 90pp; English.
The invention relates to nucleic acid comprising pederin biosynthetic
gene cluster containing pederin gene units which are either coding units
for individual enzymes or for one or several polyketide synthase or

CC nonribosomal peptide synthetase modules. The nucleic acid of the
CC invention is useful in the preparation of a modified pederin biosynthesis
CC gene cluster, or in the preparation of a modified pederin molecule. The
CC modified pederin molecules are useful as alternative antitumour agents
CC and are even more potent antitumour agents than the original pederin. The
CC present sequence is Pedf mixed type I polyketide synthase/nonribosomal
CC peptide synthase (PKS/NRPS) encoded by pederin biosynthetic gene cluster
CC from Paederus fuscipes beetle. The polyketide comprises module 1 PKS (KS
CC -ACP), module 2 NRPS (C-A-T), module 3 PKS (KS-KR-ACP), module 4 PKS (KS-
CC KR-MT-ACP), module 5 PKS (KS-KR-DH-ACP), module 6 PKS (KS-KR-ACP),
CC module 7 incomplete PKS (KS-DH), where KS is ketosynthase domain, ACP is
CC acyl carrier protein domain, C is NRPS condensation domain, A is NRPS
CC adenylation domain, T is NRPS thiolation domain, KR is ketoreductase
CC domain, MT is methyltransferase and DH is dehydratase domain
XX
SQ Sequence 8601 AA;

Query Match 22.0%; Score 539; DB 6; Length 8601;
Best Local Similarity 2.7%; Pred. NO. 0.0011;
Matches 225; Conservative 82; Mismatches 133; Indels 7873; Gaps 73;
QY 1 MGRKEMVVRDVPKMFVLIS-----ISFLVSVFINCKVM 33
Db 235 VGANLLLR--PEPFVLISSEQLSESASVHSFGAQAQHLRAEGVCSILL-----KPL 286
QY 34 SKAL-----
Db 287 TKALADGDPYASIKHSVNFNGGGASIAAPNVSHVDLIKSCYQARVDPQVRYIEA 346
QY 38 -----YNR-----PWRG-----
Db 347 QGMGNVLADLVEMQAFNRALTDIARQQRVSUPLPGNCLISTLKPMGHMESASALGALFKV 406
QY 45 -----
Db 407 IRSLHTRTHKIAHTQYHPDMYQGPFCATAGETVAMPQMEGLRAGLHCYMGGVNAH 466
QY 45 -----LVLS-----
Db 467 LLVEESVAGYDDSELGTVSSLLEHLVILVLSAKTSESLMMARRLQQLQKADAVPALRD 526
QY 49 -----
Db 527 IAYTLQVGRDAFEHRLALVDSQQQLIEGLCYLEERQPSQGEVAYOGQVASEQSULPF 586
QY 49 -----
Db 587 TEDDLAAVRCWAGAVLWVPVGPCKPRVRRLPAYFPKRYWVDSAVVEAERAPNSK 646
QY 49 -----KIGKY-----
Db 647 APASMLSGERSIGDYLRAGLGEVLQVPVERIDPQOHLVDLGVDSIVAMKLENLARAEGI 706
QY 54 -----
Db 707 PVRGRDLLQYSTVQALSRLHQAQLDRDQSVESGVEDEBPRLMASRRCSLSEQKGLWVLQ 766
QY 54 -----
Db 767 QLASEMTAYNIPLCVRIAQVLDITALEAFALLQYPLTTSVFFVQDNGELFRECCHVAAA 826
QY 54 -----KLDQ-----LKLEILR-----
Db 827 LPFWQETNTLDQAQVRMLKCLAKQPFLEKEKGLVRLHLVLSGCEHSHYLLCVHHIVFD 886
QY 65 -----
Db 887 GGSFLPVFGLLQTYQLISQQTFLAKSTRTGEYADFVLWEQRMLASAEGQRHAYWKQ 946
QY 65 -----QLETTISTKY-NVSKO-----
Db 947 LSGELPVLSLFTDNPRDAGQRTGTGTGFDLVNLSRKIRNPAKQQLNLSTLFLALFKL 1006

QY	80	-----	79
Db	1007	LLHRYSGQSELIIGMPEQSRSERFEGVVGYFVNMLPIRSRGVSKPLAEFARDLQLSMA	1066
QY	80	-----	79
Db	1067	DAMDHAVYPPVPMVRDLGRAPAEIDLAPIQVAPEYQNVFSAQDLRLFNQSYRESIGVTFL	1126
QY	80	-----	PV 81
Db	1127	EEFVQGEYELALEVREGEIDFALNLKFNPTLYRMATIARMABHLLILAHAIDAPLSPC	1186
QY	82	KNLTWNT-----	92
Db	1187	RELTMUSERERHLLHEWNATTEPYSCFHOLFQKQARMPQIAIAIFQEQRLSYAELD	1246
QY	93	-----	92
Db	1247	ERSERLAIYQQGVQPNRIVAVCLERSLDMVALIGIARSGNAWLPLDPNYPDDRLRFM	1306
QY	93	-----	92
Db	1307	LSDSQAQLLTBEGLRDKTAAIVSQAVGERLQIVAMDGHWPETIERQARTSELQMRDDPN	1366
QY	93	-----	92
Db	1367	LAYVIYTSGSTGIPKGVMIHRSLVNFYSLMNLPRGLRANDRLAVTTYCFDIAGLELIV	1426
QY	93	-----	92
Db	1427	PLLCGACCICATDKLNDSEALQGEIERLOPTVMQATPWTLLFHGWNRRQGVKILCG	1486
QY	93	-----	YYILAG 98
Db	1487	GBPLPALQRPAESASQTNLFGPTTETIWSVSRDLTKDSVDIGTPIANTRYVILNG	1546
QY	99	-----	98
Db	1547	DDQLVPVIGVGBELCTAGDGLARGYLGNPQLTAKFTIANPFCNRLRTGDLARWEDGV	1606
QY	99	-----	98
Db	1607	LEHLGRDQOVKVRGRIEISDIETWLNHRPSVAQSLVVGHEQAGMMLLVAYVVDSEWA	1666
QY	99	-----	108
Db	1667	SVSSTELRNYLAHLPEYNVPAFFRALSNMPLMNGKVDKALSARELVAESESSEGRLP	1726
QY	109	-----	108
Db	1727	CSDIEREVLDIWRSLLAVEGIGSVGFVEVGGNSILSVMLAQOISEAFGIRPAATDLFKY	1786
QY	109	-----	108
Db	1787	PTIRDISLLIGETRERSETKTGTMAAGDGGKADSVLQAQRGRQSVTGYDYYQDSLAI	1846
QY	109	-----	109
Db	1847	GISCNWPARTLRQFWENLRQKESSTRLSERELRAGVPEELIRHPDPVPMQYSMEGKE	1906
QY	110	FD-----	FYSTQL 117
Db	1907	LPDPPDFNLAKNALFMDPQYRVLQQAQWAIEDAGYVAQDIPETAVFMSASNNFYKILL	1966
QY	118	R-----	118
Db	1967	HSAGAVETDEYAAMTAGGGTIPTWISYQLGFKGSPFAVHNSCSSLVGLYLASQCURL	2026
QY	119	KPAKY-----	123
Db	2027	KEAKYALVGGATLFPVAGTGHLYTTPDMNLSSDGHCKAFDADADGLVGEGAVVLMVRKAL	2086
QY	124	-----	123
Db	2087	DAIRGDDPIYALIRGVAVNNDSKVKGFYAPSVNGQAAVIQKALDITGVDPOSVAYVEAH	2146
QY	124	-----	133
Db	2147	GTGTRLGDPVEIMALNEVYRRYTEKQPCRICSVKPNIGHLDVTAGLAGMLKVVLSEKHA	2206
QY	134	-----	152
Db	2207	EFPPSINTYREPNPAIDFTSSPPEVVVTQTPMPAGNEPRRAALSSFGIGGTNTHAILEEYV	2266
QY	153	-----	152
Db	2267	ARTDADRWNEDNOGVPLPEQVVVLSAKTORLOASVVVLYELLRAQATTEQLDQLAVT	2326
QY	153	-----	152
Db	2327	LQVGRQAMDWRVAFLVKDLHDLSEKLERFLOGDSLQDCFCQGRVATSVMDAAATPLPVAQ	2386
QY	153	-----	152
Db	2387	DREQAAIAKAWVTGRLVDWKELPRRGTPHRISLPTYPFAEERYVWEMPELPCRSESEQT	2446
QY	153	-----	152
Db	2447	KEWIEGOAERTLLVVHPLWQAHAVVVRERPLIFTEHLVWLCGFDVSLVRALTRCLPEGYR	2506
QY	153	-----	155
Db	2507	IVSLTPEGRGVNAQYQSLCLQMLERLQKIGDATKTLIQLVLPOGEYSLSFGLHALLK	2566
QY	156	NVSKEN-----	161
Db	2567	TVSQENPKVAQLIRVSSGETARSLADKLIENTFAPDDSHLRVAPSRMLDQWTLRQEBT	2626
QY	162	-----	161
Db	2627	VLDMPWKEGGVYLLTGGAGGLGVLPFAEIRARVRKATLVLVQSPLSAAABAERIAALR	2686
QY	162	-----	169
Db	2687	SDSITVICROADISCATSCSGLIADIAEQ-CGTIDGILHTAGVVRDAPFLNKSAAEFQBV	2745
QY	170	-----	169
Db	2746	LAAGVAGTVNLDRAQALGLDFFLLFSSAAAAGNAGQADYCAANAFDAYAYERNQVA	2805
QY	170	-----	169
Db	2806	AGQCRGHTLSVGMPLWRDGMRLNEEAQOAMRYTTGLVPMDSRSGIRGLYRSLAARLGT	2865
QY	170	-----	169
Db	2866	LVLEGDATAIGSLLANGTARSVSELGVPAANGNDLDELKDKTIYQLKLLAQVIGRAVE	2925
QY	170	-----	169
Db	2926	RIESCEPMDRYGLDSIAITQLNRKLEEQFGLSKTLFYQYQTVAEALAEYVLNKTVSCRA	2985
QY	170	-----	169
Db	2986	WTGLRDESVLVADAARRGLPLPETAPVVERNVLPGVNAVQEPPIAIGLSGRYPQAEETLEE	3045
QY	170	-----	181
Db	3046	FWENLQAKDCVSEIPEDRWRLENPFPHDPKPAVQAQKSYSKWGGFIEGFAEFDPLFFNI	3105
QY	182	PR-----	WN 185
Db	3106	SPREALMDPOERLPLQCAWHVLEADAGYTROSLOQGHKGVGVGVTIKTGTGFDLYGPELWH	3165
QY	186	-----	185

Db 3166 RGERLPHTSFSSVANRVSYCLNLKPSMPIDTMCSSSLTAIHEACQHLRQGDCAIUVG 3225
QY 186 -TKLYVGPT----- 193
Db 3226 GVMYVHPSTYVGLCSAYMLSDGQCRSPGQGNFVPGEGIGAVLLKPLARAQEDDLI 3285
QY 194 -----KWNDSQTIYFL----- 205
Db 3286 HAVIRSSVNHGRTNGYTPVPNPAQELIGDCLKKAGVDARSIGYIEAHGTGTGELGDI 3345
QY 206 -----GLTALL----- 212
Db 3346 EYNGLAQAFQAGEHSRCLGSVKSNLGHLFAAGMAGLTKVILQMRHGQIVPSLHAQV 3405
QY 213 ----- 212
Db 3406 LNPNDIFAATPTVPQOLVEMWRTILQESGRSRELPRRAGLSSFGAGGSNAHLILEEYIA 3465
QY 213 ----- 212
Db 3466 PBPAPRPFGEPTAAVILLSAKTPECLRRVVSLLAFIESELTRTVPDQTLFDIAYTL 3525
QY 213 ----- 212
Db 3526 QVGREALDERGLNAVLSQELSRQLAAFLGEEAEQPLLYGRVQRNKKDALQALANDEEFO 3585
QY 213 ----- 212
Db 3586 ETVDKWLARRKYSKLLKFWVTGLSVDWTRLYSVDLPRRLPYVPVRQRYWLDAYSLEP 3645
QY 213 ----- 212
Db 3646 MYPTEQPSYVPVDAEYSGSDAGREADMLMGPVDAVVEQTEDEPPAGARIAMVGGSEA 3705
QY 213 ----- 212
Db 3706 QKRAVFEQYKALELAAGAVGAASIALRGLDHHVWFAPASQTQGWADERIIDAQRDGVL 3765
QY 213 ----- 212
Db 3766 ALFQLVKLLAEGYGVABFGMTVITTTQALATCDTERIDPTHAHVGLVGLAKEYPKWRL 3825
QY 213 ----- 212
Db 3826 RALDIDARAEVFPVGLWRLLPHTRGESRYMRGCEWLRLQRLVALNGMPVAKGRAYRRQGVY 3885
QY 213 ----- 212
Db 3886 VVIGGAGGLGWTSRMMIRDHQAQIVWLGRSAKDATVRAKLDEVADDGLAPDYWQIDARD 3945
QY 213 -----RYAQ----- 216
Db 3946 ADALRQTFQVREYQIHGVIVSTLGDYDQSVAQWSEALFREILSSKLDIGVRLSQCLR 4005
QY 217 ----- 216
Db 4006 DEALDFVFFSSWAFGRSGGMAAYSACAFTDFAQLGNELACAVKVINWYNNLGGG 4065
QY 217 -----RNC----- 219
Db 4066 TRISAALKRLVEQRGVRPIEAREGLCALAVLLDGLRQLAVTRTCOPAAIETFEAGQWLT 4125
QY 220 -----THSFY-----LVNAGRNLF----- 234
Db 4126 VKAGTHSCFANVEAYQTPMQPQESPDAARLNIWVRLLFVQLQSLGLFQETGFQGNATAI 4185
QY 235 ----- 234
Db 4186 RRQAGIVDKYRWRESNLIAEHGYRLIAGDEVARIASADEIGESSRRLWQEMRECKT 4245
QY 235 -----RVPKYINGTKL----- 245
Db 4246 RFLQEQOHTLAVLVEDCLSQLPEVLRGTRLVTDILFPNGSMEKTEGLYKNNLICDYFND 4305

QY 246 ----- 245
Db 4306 VVAGVAQAYIORLENEPNAEIRLLEVGAGTGTSTTVLPQLNLWRAFIAYAYTDLSKS 4365
QY 246 -----KNTMRKLK 253
Db 4366 PFNHARLRYGTDYPYITYRLLNIEBPLIQDDIEIGTYDILIAITNVLHATRNMENTURNAK 4425
QY 254 ----- 253
Db 4426 AALRNGIILLNEISDKTIFASVLPGLIDGNSLAEDHEWRIPGSPCLPAENWQALLQSG 4485
QY 254 -----R 254
Db 4486 FDKVSFPAQVAHDLQQIIVAQTNQVHGHGAGPVLETAADKPLPTLESAAERLVD 4545
QY 255 KQAPV-----KEQFEKKAKKTQS----- 272
Db 4546 SSVPARRODVAARVELILDSLAQALSTGREQIE-----QDIPFSDYGIDSILGVGFVQ 4599
QY 273 -----TTPYFSYTT----- 282
Db 4600 RLNDELGLSLNTTLLFDYTTVQRLAEHIVAEYGHITLDVPAALPGPELSVSEBAMDIPLPA 4659
QY 283 ----- 282
Db 4660 VQAVPSSLPRREAVVQTDGIAVIGMAGOPPGADSVDALWQNMVAGVNPVTELSLYLPYH 4719
QY 283 ----- 282
Db 4720 AYSPEKQPKSYCKWKGALQGRDCDFDPLFFNISPREASMNPHQRLIQESWKALEDAGY 4779
QY 283 ----- 282
Db 4780 APRSLDSRTGIFVCAEPSAYVHESFVGASDAIVASRLSYFLDLKGPVAVVTGCSGGV 4839
QY 283 ----- 282
Db 4840 ALHLACESIRNGETEVALAGGVFVWGQTLVGLAQTDMLSRGTCCTFDADADGMVSE 4899
QY 283 ----- 282
Db 4900 GVMVVLKRLDQALSDGDTIYGVIRASGINQDASNGITAPSGIAQQQLITVYRYAID 4959
QY 283 ----- 282
Db 4960 PRITYVEAHGTGTGLDGPVEANALVRAFRSFTESTGYCAVGSIKSHIGHTSSSSGVIGL 5019
QY 283 -----SAALN----- 287
Db 5020 ISILLCLKHHQLPGMRHFKRLNPLIEFERSPFYVNARMWPMRSGSGEPLMAALNSFGHSG 5079
QY 288 -----VTTN----- 291
Db 5080 TNVHLVVEEFVRNSDEPRVDDVSSTAQPELILLSTKDAERLSEVLNNLAHFVQAQNO 5139
QY 292 -----VTYSITTA----- 299
Db 5140 PADLERLSLADLAYTLQTGREAMEQORVALLVCDLAGLLEALSALREERPCPVSWMSGRVE 5199
QY 300 -----ARRV- 303
Db 5200 PGPSRGAETVNADQPAELLQRIPOWLAEGALDELAQWAGAPIDWCQLRRRRPPRRVH 5259
QY 304 -----STST----- 307
Db 5260 LPSYPFARERYWRSEPAVHSPVVAAGLHPLVORNTSTLDRHCFFESFGSEFFFRDHRVQ 5319
QY 308 ----- 307
Db 5320 QOPLLPAYLEWARAAQIALGNACPDVALKLSNVVWIGPLLARQPIVGTITLQAREDR 5379

QY 308 ----- 307
Db 5380 GIDYQISSVAAGQPVVHCQGIATTETETEKEAPVLDLALRSRLTQKEIGVERCYAALE 5439
QY 308 ----- 307
Db 5440 AAGVNHCPAMGLQAVSRNAEEVLATLRLPAETVGEASAVVLPAILDAALQASIALTLR 5499
QY 308 ----- 307
Db 5500 DDEVPSPETSPRPVLLPFALESRLVVPCCASMAWIRLVAVEHAGQALQRLDVIDCTK 5559
QY 308 ----- 307
Db 5560 EGEVCVALRGFTSRSLPPSGATESRASASASTLVSTEGVSRFGEBFPLRHDHGMFLPA 5619
QY 308 ----- 307
Db 5620 AVYLEMVRAPAEKGHERKITGLSHVWPVKVLLVSGEGREVRTCLTNVDRSAFLISACEQS 5679
QY 308 ----- IAYR ----- 311
Db 5680 SEGQEVTVYCOGNLLPEVMEEPGAALATBAIAYRCPVLEAKQCDRLQLQSTHGPALMSV 5739
QY 312 ----- 311
Db 5740 QQLRYSDREALLOLPDELQMGDDYGMHPSLLNGAILASVVMCLARAPRSRAGLPMDF 5799
QY 312 ----- 311
Db 5800 SLDLRVFOPFERQMOAYVRRHGSARSNGENLEKVDIDLLDSQGRCLASLEGTLVFPD 5859
QY 312 ----- PD 313
Db 5860 ANELVVAIPQWVEQALPARVAASAPLAVQAPVILAGAGEPLRRALHDSWPDALLHELPE 5919
QY 314 SSP ----- 316
Db 5920 SAFEVGDGLRQAVVEVFGMCRLLPYKGAALOPLLVLLPEAREVTPQALLGGLSGLLK 5979
QY 317 ----- MK ----- 318
Db 5980 TVRLEHPRITARIISYPVDDTVTAGMWKVLAAEIASPEGDVEIRYDQARRHKVLHEIT 6039
QY 319 ----- 318
Db 6040 LSAGEHGDLSLFRPDDVVMLTGLGGIGRQIARVILGVERRVRLALSRSALDDKGERFLQE 6099
QY 319 ----- 318
Db 6100 LRREGAVSVLYRVDVADADAVGRALLAIEQEHGGLTGIIHSAGIIADDVLSKTTAQFEQ 6159
QY 319 ----- 318
Db 6160 VLKPKVSGVNVLDAAATANRSLRYLLVFSSTAGVGNMGQADYAAANGFLDSFAHYREALV 6219
QY 319 ----- 318
Db 6220 RQGLRSKLSLNLWPLWREGGMQRHGEALMQOATGMLAMESAQFEALEAGLRSAQAQ 6279
QY 319 ----- 318
Db 6280 ILVAFGEVPSIRNRLTFRMDAPEPPAPSVMEMDRAPGEVGEQETQQLVRSVEAELIRI 6339
QY 319 ----- SIMAT ----- QLRDLA ----- 329
Db 6340 VAFVQIPAEKINVRDRISAYGPDISISFTEFANALNKAYKLSIMPPLFFEFIASLADLAGH 6399
QY 330 ----- 329
Db 6400 LLTQHRPALLEKHANDVEKPSCHSGVAAQIPIPTMPKSECIATLPLLVGSIPEPQAD 6459
QY 330 ----- 329

Db 6460 LEAVAVIMAGKFPFGCEDLEDFTWCQSCODLISEVPEORMDWRRFYGDPHQBPQTKIK 6519
QY 330 ----- TW ----- 331
Db 6520 WGGFVADACFDARFFGISPVEAEVMDPOLRLPLETVMAALEDAGYPAGLAGSRTGVPA 6579
QY 332 ----- 331
Db 6580 GVATADYKOLLIEARARGLVQTPSEPPFPMIANRISYWFNFNGPSEBDAITACSSSLIAVH 6639
QY 332 ----- 331
Db 6640 KAIESRLGSCWALAGGVNLGSPRITTIASSOAGMLSEBGRCMWTFDERANGVYRSEGVA 6699
QY 332 ----- VYT ----- T 335
Db 6700 ILLKPLRKAIAADNDRIHGLIRGSGENHGRSASTAPNGNAQKRLLLVDIYSRADIDPRT 6759
QY 336 LRY ----- 338
Db 6760 ISYIEAHGTGTVLGDDPEVNGLKAAPQELYQSRGLDVPEQPHCGLSNVKANVGHLEAAAG 6819
QY 339 ----- RQNPFC ----- PSNRT 351
Db 6820 AVGIKVLMLKHKRIPGNPHLRPNPYLQLEGTPFYLRETLDMPQPTDVRGNPLARRA 6879
QY 352 AVSEP ----- MNTHVLIR ----- NETPV ----- 370
Db 6880 GVSSFGVGSNAHVILEEYQBPQGWGSEPAYPALIVLSAKDEERLVCVAQRLLRPIRD 6939
QY 371 ----- 370
Db 6940 YGSELVLDHIAVTQVGREAMPRLALAVTSLAQLADRLQTLWEQPTQTEGVQOGLVTOE 6999
QY 371 ----- TIYG ----- TLDMSLYNET ----- 386
Db 7000 ABEQFDTVLGDDEDAARAVERWEKQYKELDAWTRGWAIDWNVLYCTDTPRRRIGLPTY 7059
QY 387 ----- 386
Db 7060 PFARRRYVASVPAEDRGNSTLSEBPEQRSKSDLLTFEYWAEPVLAAPATDRVKTIL 7119
QY 387 ----- MFVENKTASDNKTTPSPSMGFQRTFIDPL ----- 417
Db 7120 LCICSDPHEHORIAEQXDSRDPGVQLIFIEQDA ----- PAEPDEARQR ----- IDPLQPS 7170
QY 418 ----- 417
Db 7171 SYSRALTTTIKALGRVDALLYLWPCEDRRWISNVLPVLHLLQALYETGLRPRKLLSGEY 7230
QY 418 ----- WDLDSL ----- LFLDE 429
Db 7231 ADALERCHLDSWAFERSLGVMVMPETQVAVFRERAADTGESSPTWDLVLEVLFAEK 7290
QY 430 IRN ----- 432
Db 7291 LRSACYRQVRHVPLIRPLAWQGSASPFKQGVVLTGCGGGLGMIVAEHLATVYAARL 7350
QY 433 ----- 432
Db 7351 VLSGRSSSLAEKVELLQARGAQLVYGADVTDVHAMQEWVDQARRHPSPLNGVLHIAGL 7410
QY 433 ----- 432
Db 7411 NGTAEVLKABADAFORVLDKATIGTSQVLDQVLRRESLDFICYFCSSAIIIGDFGSDYAL 7470
QY 433 ----- 432
Db 7471 GNRFOQALYRAQWVSSALSCKTILANWPLWQDGLGVGDAEQTRFYLOSGGQSLCS 7530
QY 433 ----- 432

Db 7531 QEALALLEQLTQDRAQLVWAGQPDRLRLRWVNBQBLEAATVTPBPVRAKAVARAEL 7590
QY 433 -----FSL-- 435
||
Db 7591 GGGDLQCLLDLTKTKICELIGTQVNELEHANLVDFGDFSISLAESRVLRSFYSLDI 7650
QY 436 ----- 435
Db 7651 SPSVPFSTLNLTAYFLAEHRQTLEGFYQQPQAPGPEHAPVPTIEVAQSVVPVTALL 7710
QY 436 ----- 435
Db 7711 PTGTSIGSASQODEIAIIGLSRPPQARTIEELWRILEOGRDAIQEVPIDRFWRSY 7770
QY 436 ----- 435
Db 7771 SPSQEMSKSNKGGCIPGIAEFDFLFFBISLEAERMDPRQHLMOEAWLALEDAGYGP 7830
QY 436 ----- 435
Db 7831 EQLECNKISMFGVEBGCDYQRRLTQQTSLTSMHNGILASRLAYFLNLKGPVMAINTACS 7890
QY 436 ----- 435
Db 7891 SALVAVHQASLRHGECDTAIAAGYNLLVAPEAYVGMTQAGMLSPDGKCYVFDKXANGL 7950
QY 436 ----- 435
Db 7951 VPGEAVVVLKRLSKALADGDPKALIRGSGINVDKNGITAPSGASQTELLBGIYRQ 8010
QY 436 ----- 435
Db 8011 CALQPDISIVTHGTGLGDPIEINALVDYFKGTDKQGFALTSIKSNLGHTPAAG 8070
QY 436 -----RSPTYVN----- 442
|||
Db 8071 LVSLISLVLAIRHRTIPSSLHCEQKNDYIRWQESPFPYVNTKXKHECALGPRIQAVSAP 8130
QY 443 -----LTPPEH----- 448
||
Db 8131 GMSGTNAHVVQEHQPAEFSRWSSTAAPYLVWVSAKTETTLQEQIRQWEDYLSRHPDLDF 8190
QY 449 -----RRAV----- 452
|||
Db 8191 EAVSYTLKGRHHFKYRCAIVAKDLSQLQALRQALDRQTQANLCMGCVDRDFSGOKAIR 8250
QY 453 ----- 452
Db 8251 DFTASLAAQGEALURDKPDDYRDNLIALADFYCGYEVAHVHLFAGRPQRSLPGYPFARE 8310
QY 453 -----NLST----- 456
|||
Db 8311 HYWIDPSAAHRAELSERSDTQNLPLQRLNSTLSEQRVYSAFAKGFERMLRIAHQBEL 8370
QY 457 ----- 456
Db 8371 LIPTLFLEMARLAAQSLMDPVRLKNMVMACPLYQQGSDYELFSLHEKDSLLYTV 8430
QY 457 ----- 456
Db 8431 EMQGEVPVCGHFGIEIDSTEALQLPVVPDIARLRLASLFPVVPVALGAPEMGVARISDVQA 8490
QY 457 -----SNSLW-----WWL 464
|||
Db 8491 DATTLATLNVADGEDRSWVHFPLLINAGWWL 8523
|||

RESULT 11
ABP76678
ID ABP76678 standard; protein; 19938 AA.
XX
AC ABP76678;
XX

DT 26-FEB-2003 (first entry)
XX Streptomyces viridochromogenes Avi gene cluster polypeptide frame 2.
DE
XX
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
OS Streptomyces viridochromogenes.
XX
PN WO200268436-A1.
XX
PD 06-SEP-2002.
XX
PF 24-AUG-2001; 2001WO-EP009815.
XX
PR 25-FEB-2001; 2001DE-01009166.
XX (COMB-) COMBINATURE BIOPHARM AG.
PA
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX
DR WPI; 2003-018650/01.
DR N-PSDB; ABZ37515.
XX
PT New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes.
XX
PS Example 1; Page 68-301; 319pp; German.
CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37515-
CC ABZ37516)
XX
SQ Sequence 19938 AA;
Query Match 22.0%; Score 538; DB 6; Length 19938;
Best Local Similarity 1.1%; Pred. No. 0.0098;
Matches 223; Conservative 92; Mismatches 141; Indels 19088; Gaps 71;
QY 2 GRKEMMVRD----- 10
|||
Db 359 GRVERAGRDGWAGLEDRAGQPSRPGPLPPRAKPPSPRAAPLPPGMSPLPPRCIPSSRPA 418
QY 11 ----- 10
Db 419 PPPPPRATPRPRAACDNGVGVACRASPAGRVLCGSTGPTYXVGLCAESPIRHAVGRTL 478
QY 11 ----- 10
Db 479 AMCSIRHRGRVHVADWNIWIGVCVCICTEDRRRTGFPQSPQXSPLTYGKSQVVAHDRNR 538
QY 11 ----- 10
Db 539 SAECRQVDPVQRPDQERRAGQLPVRHDAQRRCRCRPRRTPHQVGRDLLPAHPGDRR 598
QY 11 ----- 10
Db 599 LRRHRHRARRLURGGGPGQQVPREHPXVRRDLPGHPRLOGRRRRRCROGLAQGRHDDQ 658
QY 11 ----- 10
Db 659 HRADPGPDDREGPAPPEGLAQEGPRAEGEGRGQDPEGGRHPLRPRHRPCHRAQ 718
QY 11 ----- 10
Db 719 RAPAPAPAHQALPLRLQRRGRADRLQERAAARPGRPRRHLPRQAGGRPRARRR 778
QY 11 -----VPM----- 14

Db 779 RGRTPPVRRPGRARPRRRLQHPADLPDGRPEGIPRLDHOEGRRHPRGRRRHPH 838
QY 15 -----:|:|----- 14
Db 839 RLPGLHOGGHLHHPGGDLRRRCPREBEGAYGGGLCDAGRGRVPLQRVGAIPPR 898
QY 15 -----:|:|----- 14
Db 899 RCFQTRRSEGVDSRGRPLAVSRAGMVLGKLIHQSLPCRMRSVVYXVPLEGRGRIMS 958
QY 15 -----FVLSI-----SFL----- 24
Db 959 INRERSKALFPLIKKNDNHEAIEIVSKHNAVLSAEDYAALREGSYLLRSPANARKL 1018
QY 25 -----:|:|-----:|:|----- 24
Db 1019 LKAYENALAHVNVSEBELIDPDSADAGSNAAGLSSRIIRAGMTTRPGSRTTARCSFSTS 1078
QY 25 -----:|:|-----:|:|----- 24
Db 1079 SSRTGATPSRGSANPSRXTTCRGRGSGTNTTASCTWLRTRRSXSLPGTTTDCGG 1138
QY 25 -----:|:|-----:|:|----- 24
Db 1139 PPVLGWCWGPLEPDVSAGSSVNVPLQRLGADPPRLVEACAGQGPTHRSRAGSPFC 1198
QY 25 -----VSTINCKVMSKALYNRP-----WRGLVLSKIGKYK----- 54
Db 1199 LRQTGSCSVITCRSGSRALSRAPARTSMKG---SSGOPYRXGSRRRRCRSGPGXSRR 1255
QY 55 -----:|:|-----:|:|----- 54
Db 1256 RRSRTWPCAGRCRCVGTWRVGLRRSARRXGVSAFTRSGGPTRHGXSWLRRRXGDGS 1315
QY 55 -----:|:|-----:|:|----- 54
Db 1316 VRRGATRAAGDRDFSGRSGAGSPCSRSARSRSRSPRSAGTASATSPGTDYGRSR 1375
QY 55 -----:|:|-----:|:|----- 54
Db 1376 WBRRCRRRGLRRLPAGTRCGLRPGVPRARSSRVPSPRCRCRCCTCGRRPPARTGSG 1435
QY 55 -----:|:|-----:|:|----- 54
Db 1436 SLGRGRPGRXXSRXSGRAPSSVRAPFPDAGVVSVOQGDQGGAAWVCYLFEDLDGGQP 1495
QY 55 -----LDQK-----LEIL-----:|:|----- 63
Db 1496 ABAAQGRPADDRKDCVTTDDQLRAGGVVVAVELDDEGLDVGGRAGPGGHDARGGE 1555
QY 64 -----:|:|-----:|:|----- 63
Db 1556 PPAERRAVVGPRAVLGCPVFEAGACQVDGAPGGVERLVLGRRADVPPTVLGAGRIDPVR 1615
QY 64 -----:|:|-----:|:|----- 63
Db 1616 RHGSLAGLKAGRAVLXVPGAEDLLRMARSGERTPAARRITSRNARFOQWALLGNR 1675
QY 64 -----ROLETTISTK-----:|:|----- 73
Db 1676 NKRTRAGEFLVMGVRPISLAVEHGWFTLLYDQRELSKWARELLRTVTEQIAMAPDL 1735
QY 74 -----:|:|-----:|:|----- 73
Db 1736 LMELGKNEAPPEVAVVEMPADDLRIIPVREDFLGVLFDRPTSPGNIGSIIRSDALGA 1795
QY 74 -----:|:|-----:|:|----- 73
Db 1796 HGLIVAGHAADVDPKSVRSSTGSLFSLPAVRVPSPEVMDVWEARRAAGTPIVLVGTDE 1855
QY 74 -----YVVS-----:|:|----- 77

Db 1856 HGDCDVFDDFTQPTLLLLIGNETAGLSNAWRTLCDYTVSIPMAGSASSILNAANAATALLY 1915
QY 78 -----:|:|----- 77
Db 1916 EAVRQRISGRTATTXAAAGSGCTWPAPGSGAPRSAPRAARSRLPDGPAMQGRGSPASEPR 1975
QY 78 -----KQPVKN-----:|:|----- 83
Db 1976 RHPRTRTXIPLRAAAPAGSPTRAFRAGXPXTAPCRPRAGXPAPAXXRPVPSRXHVRTA 2035
QY 84 -----:|:|-----:|:|----- 83
Db 2036 PAAGRRRFRSRSRGGDVLEAGAGLVPFPFPADPNPQPGLSGGHLDLRLPPFPVAA 2095
QY 84 -----LTWTEFP---QYYI-----:|:|----- 95
Db 2096 DVLQVLVPLRTELPMWYQLGHSETGEBLPGVGQARQLQDEVAERGVEPSILDVRG 2155
QY 96 -----:|:|-----:|:|----- 95
Db 2156 ADGAVVGGDTAQVAVRNRLSGAGPWETSRPVSARPTDTAGPRTXTQDVFPSARPAPPE 2215
QY 96 -----:|:|-----:|:|----- 95
Db 2216 DAAGSARPSAKARRXPGHSPHLPHPHRRQRSGXIPPANEDSRRLAPGLPFRPPKDTTP 2275
QY 96 -----LAGPI-----:|:|----- 100
Db 2276 VATTGAVSLGRKNWNTGENTGDKTIQPGPLLRGLPPARRLPGPVRPSCGRSSAAPG 2335
QY 101 -----:|:|-----:|:|----- 100
Db 2336 PLREPPSGYCGRXXRSPHARAAPTRASAGPATGCAGTSPQCVRSTAGRPARHART 2395
QY 101 -----:|:|-----:|:|----- 100
Db 2396 RRGRRTAHADRRRRCGWRTAAHRPPPSGAPAGGRPARPPDRRACRNARSAAHVATARAR 2455
QY 101 -----:|:|-----:|:|----- 100
Db 2456 SARPARAGDSPPRAGRPHRRSPAAASAGAAAPASGGNRCRSTGTASPDRRRG 2515
QY 101 -----:|:|-----:|:|----- 100
Db 2516 GRGPAGRSRSPGGRGRGNPARSGARHRRDRNCRSAPRRGTHSPVHSPVLNRLRRSP 2575
QY 101 -----:|:|-----:|:|----- 100
Db 2576 EAIIVPRPRPGKLPAAAGHTGHRRLGCRSSQVKGRRGRHPNLXGFPECAPTYDCRTP 2635
QY 101 -----:|:|-----:|:|----- 100
Db 2636 MPXLRGRAGHVHPGVLSRNTSAGDRRGLHRLARGRLPGGRGRGRAGRPDHRXPSTS 2695
QY 101 -----:|:|-----:|:|----- 100
Db 2696 GPTCGHPSRRDRCRRCRGRVAGAPRHLPSRGPDRRPFGLGHTGRRBAGQRGHDORA 2755
QY 101 -----:|:|-----:|:|----- 100
Db 2756 GSGARRRGAGRVRLHRRGALRGRSGPDRGHAAQGGALRHGQVLURGEVHRPVQPAARN 2815
QY 101 -----:|:|-----:|:|----- 100
Db 2816 RAQRVAARQVRPAEPGPGXGGRHRLLAGLRGSAHGVRRRFPADPLXLRVRRRRRGVR 2875
QY 101 -----:|:|-----:|:|----- 100
Db 2876 RAVHRRPGVWNIGHQGEHGGAGPRHRRRLRARPSPFRAPPAGRDPAQHGRHSCRR 2935
QY 101 -----:|:|-----:|:|----- 104
Db 2936 RSGLDVRVAGERHRRHRLGLGPFRFARPAAGMTNLRSSRSPWPTPIAVALPYTONSSGV 2995

QY 105 ----- 104
Db 2996 QRDGAVCRVSSLRSPAADWCRRASERTLPVGAQEAEGEPAGVDIVDDRVLGQQAAR 3055
QY 105 ----- 104
Db 3056 AHPPSSAEXRCPHRTAGXRCGSGFGRXPGGARSRPPCRPRSCXPRGRGASRR 3115
QY 105 ----- 104
Db 3116 GGTAPRTGTAAGALAAVAHAHRPRAPRRGTGRRRPRRRPPAARRRPTARAGRPIPA 3175
QY 105 ----- 104
Db 3176 APCVGTPTALAWPGSASSPRPSPGWRPSRQVSPAGPAARIRHRSTGRGCAADAGGR 3235
QY 105 ----- 104
Db 3236 SPDGRHRPPPLDAARGCGFQSLSCPQLQYGCVPAGALASVSRHSSYELLKKCAKE 3295
QY 105 ----- 104
Db 3296 SAAPTALALLVSKRRPTRTTGQVRDVGSRAPHSGEXPGADLSRRRLPPGGRWNLSALANT 3355
QY 105 ----- 104
Db 3356 FSGTAPVLSRGRGTRRDADGGAGHPDRRRGARI PVRGPSCPRDAGPARDGARRETSGR 3415
QY 105 ----- 104
Db 3416 QAGRQPPGDRPGGPHVPGRTAGRQPPCGAARPEPVRPARRGVTGRRRGPGHPYR 3475
QY 105 ----- 104
Db 3476 APLGLRLVAGORGAVAHPELDRDRHRTRAGDLAGRGCLLPALRDPAPRPAR 3535
QY 105 ----- 104
Db 3536 SDHVHGRHEHLWRACPEPVHLPAEPGRAGDPDHGATPLDPAGHRVPPLLALARPAGEREL 3595
QY 105 ----- 104
Db 3596 PVLDDGGFADPXGGRSAARRRRERFLRPRPEPRPNRPVRRAAVHRSAAALQPGGAR 3655
QY 105 ----- 104
Db 3656 AGRPPPAEARGAGAGPNAVITPGLSPQGTADLGTGRGRI GMMVGSGLCIADVSGKPVGEGNQ 3715
QY 105 ----- 104
Db 3716 SGLKTGDPPELRQYLGHMSLDGALGDPEQRGDLTVGVAVAQOEEDLELAVGEXMYLAEPF 3775
QY 105 ----- 104
Db 3776 QFCELVLVKSIVARITGIAAGIWOAFVXAGCPMPXEVIRRIEYQAGRHHXTMT 3835
QY 105 ----- 108
Db 3836 XGRYHPCDEWVLDITHLDHLIASTAGLRLGAACREQQPPRRGGKGGPCPCVCAQR 3895
QY 109 ----- 108
Db 3896 RKERGYSAWRHAQAPVMHLEESGIPAAGLPXNGARHLWSAGHARPVINGSLRDIKKP 3955
QY 109 ----- 108
Db 3956 V\$INFRIGDMRVLSSATKAPATARDSGSIDAFLQNPGGPHLGSDQSQRVGLDEGGAVG 4015
QY 109 ----- 108
Db 4016 HPEALHHLADVLDPLAAEVELAGDAGGALTGCGDAFDLDLVQAGXGRGGSHLVLPGRV 4075

QY 109 ----- 108
Db 4076 LRGVESAGAGPGTGGXLLCTQHSFTSSSGLSFRCSRFLDIGLFEQMGVRVSGACGLTGL 4135
QY 109 ----- 108
Db 4136 HDHGGPQEAATARXLRRSGVXRPHPALWRARSGRETAGNRHLAGRKTRGTSRGGVGRDP 4195
QY 109 ----- 108
Db 4196 NHGRRRQDXKFPYTCPSYPRRLRICGVERVETAXSMCCSRPVVTWIRCGTGAEPSPGSP 4255
QY 109 ----- 108
Db 4256 RSPSNISICCTCTRERCAACIIITRLSSTCSSRARACXSPTTPKTPTAKRNSSTSRAG 4315
QY 109 ----- 110
Db 4316 SAPGRRPGSTPTSPRPXRSSPCSPSGTNAIRWFRSSRCRTRWSAREGRGRRRGVR 4375
QY 111 ----- 110
Db 4376 RHGRHPGTGFRGHSRDACQLRGTPXRPLRRAGQRGLPLPALLGPAAPRRRPRDRGRDP 4435
QY 111 ----- 110
Db 4436 GAPSRLGVPLRPDOHHLRAHPAGHPLRPEPGRGRGALRGPSGRPPGPHVRQQHQGRSH 4495
QY 111 ----- 110
Db 4496 RHPRRAABLRARRVAPELRAGRVVRRGLGPPRRRGA VGGGADGQPGAGDQGRGFAPR 4555
QY 111 ----- 110
Db 4556 LRRRQGRPVPPGQRTLAGRRRCHRLAQGXPPSTDWIRIMSTTGHSTVIDRCRICDNTEL 4615
QY 111 ----- 110
Db 4616 LPVLDLGPQALTVFPFTRGEDVPVPLELVRCSPGGCELVQLRHTADFGLMYGBGYR 4675
QY 111 ----- 110
Db 4676 SSLNRSMADHLRGKVAATGLVDLPGDLVVDIGSNDGTLAAYPADGPRLVGVDPAAATV 4735
QY 111 ----- 110
Db 4736 FAASYPPGVVELIPDFFAVDLLGGRRAKVVT\$IAMFYDLPRPMEFMRVGRLLTDDGIWVT 4795
QY 111 ----- 113
Db 4796 EQSYLPAMLHACAYDVVCHEHLDYVGLQIEWMAERTGLKVVDABELTPVYGGSLSLVLAR 4855
QY 114 ---STQLRKPA--- 121
|: :|
Db 4856 RGSSQVNEPALARIRAGETDLPYAEFARTEESRDLLEFLTASRDKGLHTLYGASTK 4915
QY 122 ---KVVY-- 125
|: :|
Db 4916 GNVILQYCGLDETLLPCIAEVNEDKFGCYTPGTNIPIVSEBEARALEPDDQFLVLPWITRD 4975
QY 126 ----- 125
Db 4976 AMVARERDFLASGSLVFPPLPLEVVXSGGGDIARHHPRACRVPAVGAAGRAAVALA 5035
QY 126 ----- 125
Db 5036 RHHGGRVATRLRECHAGCAACRGVAAPRRPGGRRAGLSAGRPALGPATRGORD 5095
QY 126 ----- 125
Db 5096 DLROSRPFGARPRRSRGRPRPLPRPPLPRSRPPDRAGAVHPDRHGVXHPXRRGP 5155
QY 126 ----- 125

Db	5156	HRGRGNRRRLRVPRCLSGIRAVHRLRPGRRRLPAVHPLFGHHRHRLPRAGAADD	5215	Db	6236	RSQBPCAVGSVKSNIGHLEGAAGVAGVIKAILALDEDERI PASLLDGDNPENIDWAGLDI	6295
QY	126	-----	125	QY	176	-----	175
Db	5216	TRDAPVPGHLRHGRGHRHRSRLAAAGSRPRPARRPAPGRAAHPRAHRLRRHGGR	5275	Db	6296	RLATRALPWPBRPHPRRAAVSGYGGTVAHVVLBQAPTAPAPAPAPAGTLPVPSAASP	6355
QY	126	-----	125	QY	176	-----	175
Db	5276	PPRLATPGAAGAPARRRAHTDHPRPAPVRGRHRAHRHPSHLRRLPVRRHLPEDRAGP	5335	Db	6356	EALDRRAALAEVBEAGDASVGHITLAHROSPLVHRAAVAVATGRDELAAGLALATQBP	6415
QY	126	-----	125	QY	176	-----	182
Db	5336	GGPRQTAGTRHRTVAVRLGRPEDRPAATRPGRRRPVGVRRRPAGGRRRTGGGQGG	5395	Db	6416	APGLVTGAALPDAGRPPVWVFSCHGSQWAGMGRELLAEAPVFAEVIDELPVPKEEIGFSP	6475
QY	126	-----	125	QY	183	R-----	187
Db	5396	GSADVGGARRRPAHRAHELPRDTGDPHHRARTRRDTRVHHAGHRHPVRHRDRA	5455	Db	6476	RQMLLEGDHTVEDGAQTMIFAMQLGLAALMRSGVEPAAVIGHSVGEIAAAVTAGALTVT	6535
QY	126	-----	135	QY	188	-----	187
Db	5456	AARPAPRRGAHRAHPYKSEFRSLTTVRRILPGKCRSGSRXFPQBYRONADYDQKNAD	5515	Db	6536	DGARLICRRSLLLREAGRGAMVSLPFDPAERLAGNDVAVAAIASSTTTVISGDPG	6595
QY	136	-----	135	QY	188	-----	187
Db	5516	PCAPGVSUKTRGWGADSRPVHRCRRLRTAVCERRMGRGTRPLSGRRRGTRARRGRH	5575	Db	6596	EVEKVGRWTDGLVVRVSDVAFHSPHMDPLDLRAADELSAPSAPHTPLYTTALAD	6655
QY	136	-----	140	QY	188	-----	187
Db	5576	SRRRPSSGNLAGGAEVGIADGIRSAVCEHLAPGCPGLAAPTIPAAAPGGRGHA	5635	Db	6656	PRATVTADGAYWAANLNPVRLAAAITAAEDCHRAFVELSPHPVVAHSIHETLAERGVE	6715
QY	141	-----	140	QY	188	-----	193
Db	5636	RPGDQAGLQRRVQAGTRRLPAGRSEPHQADRRRRQLRNAADRPVADGPRLHRRRCAP	5695	Db	6716	DVFGFTLRRNQPEARTFRAAVGAAHCHGVSVDMSALQPNGNLEVLPPVYQHRPLWRSI	6775
QY	141	-----	147	QY	194	-----	203
Db	5696	ALVLQAARLRPALLGLOGAAGDRVAAPRXRAVPAEHHGPGCHRLQRLPYRPVRRPQR	5755	Db	6776	AGARAERGHVDVSHTLTGTPGVAGSDLWHSTLDDDSRYPGSHALNGVEIVPAAVL	6835
QY	148	MTC-----	150	QY	204	-----	203
Db	5756	LRCGDSGPDYGRGRARPGRGHRDGGRRPRLLHLSREVVQCXMAANGVPLSRSTWE	5815	Db	6836	AVTFLAAGAEGERRALQDWTMTHPVLTAGORQIQVREGEVVRLASRTVADAADPNPAW	6895
QY	151	-----	155	QY	204	-----	203
Db	5816	FGREIGHLRGPSSAVHGHTRTDXRPRRPRPATRHGPGGVQRRPSGSRIGLGRMM	5875	Db	6896	LVAEBARTAAPDLAAGLAARSLDPGEHRLPADPGLVSRRLAEVGPSTGFDMSVERLSA	6955
QY	156	NVSK-----	165	QY	204	-----	207
Db	5876	NTGDERNMPVISLDRPADSLSPVAVVGICRFPQGVNSPGFWDLLTAGRNTVGMPPD	5935	Db	6956	GLGVLHAQVLSPDASSWAPLLDAVMSIAPAFVGLPQLRMVVHVDEITVDGTPPEAATVE	7015
QY	166	-----	165	QY	208	-----	207
Db	5936	RWEYRDFGPRDAAALRTAIRSGSFLDDDIAGDAEFFGISPRAELMDPQORLMLEVAV	5995	Db	7016	VALDPRVADTVHALVTDGGRPVASLRGLRYPVVBQPAAPDTDEPGDADADVSPAGLS	7075
QY	166	-----	165	QY	208	-----	207
Db	5996	QALEHAGIPPHTLAGTDTGPAGVCTDYDAGRLDLPNIDAWTGIGAACAVSNRVSJA	6055	Db	7076	PEELASRVLDEVREQIAQEMRLAPTALHVRPLVEQGLDSVMTVVVRRLEKRLGRDPA	7135
QY	166	-----	165	QY	208	-----	207
Db	6056	LDLRGPSLSIDTACSASLVALHTAAQSLRGECTVALAGVNLVSPQOTIALGTAGALA	6115	Db	7136	NIFWKLPTISDIVHLTERLTHEPTADGHASXTHXLRNSCQGTAGXKGRASRPGNLGKT	7195
QY	166	-----	175	QY	208	-----	207
Db	6116	PDGRSKPFAASAGRYAASADGYRGEGCVLIVIKLLTDAVDRDRLVLRGSAFNQD	6175	Db	7196	XKTRQYDRCAETPDPAIWSSLETTXSDGNDKRVWDKALILVGGVSRRLRPITHTSAKQL	7255
QY	176	-----	175	QY	208	-----	207
Db	6176	GRTNGIMAPCGQAQEHVMERRALTAAGVAADTVDFVEAHGTGTRLGDPMWIGAIAVYGRD	6235	Db	7256	VPVANRTVLSYVLDISIKEAGISEVGIVVGATAAETQASVGDGAEGFGLDVTYIQDAPRGL	7315
QY	176	-----	175	QY	208	-----	207
				Db	7316	ADAVLSRDFLGDDDFVMYLGDNVYVDGIVDFVDFRDKPAAQWLAARVADRRRGVAE	7375

QY 208 ----- 207
Db 7376 LDENGRTAVVEKPDQPRSDLAIVGYAFSSAVHEAVAGIRPSWNELEITDALQWLADH 7435
QY 208 ----- 207
Db 7436 GHEVGSTVIDRYWKDVASVTDVLEMMRHLESLETRIDGEVDEASBELVGRVVVEAGAKIS 7495
QY 208 ----- 207
Db 7496 GSRIVGPAVIGPTVVQDSYLGFTTSIEGNCRAVHSEIQYSIVLRCASIEGVARVEDSVI 7555
QY 208 ----- 207
Db 7556 GRETEVNAAPTSTAHLVLGDHSRVQVGXGTGLRRRTTRRXRLTGTCETAARVNGDG 7615
QY 208 ----- 207
Db 7616 TSWSVERASSAPSTCANWSGTAIRRGPCWTSXPTPGTWTWSRSPAGTSPCTATSAMP 7675
QY 208 ----- 207
Db 7676 GCSRWSPATIWNSTSRPNRTSTGRSPTRRRSYVPMCRASRHSRCAMRPARPGSCRSP 7735
QY 208 ----- 207
Db 7736 TRCTAASRAPGTRTRERSPTTRPRRAATWSPSPTPAPMGCRCXAXEGAATTTGPISS 7795
QY 208 ----- 207
Db 7796 RRRSFRCSPPGCWTATASRCTGMAATSGTGCTPTTCGASGSPNAGCRGRTTSRAPS 7855
QY 208 ----- 207
Db 7856 XPIWNXNNCWTWPAPAGTSSVYPTGRTDTGATRCATPASARSATHPVCRSSKASPTFC 7915
QY 208 ----- 207
Db 7916 AGTPRTGTGGSXTSARCPTAMLWPCRQHPXLGESPPAGFPFADRLICFALFSRSP 7975
QY 208 ----- 207
Db 7976 GALSDDRSLRLACGGSAQPRAPIGIPRRPQRRHRACPRGRAPVHRVGDRCAGRO 8035
QY 208 ----- 207
Db 8036 ALRRAAIRPLRLGRVVGGSQSLGEEPQLRHQDQREPHQRGGAANRLVLLFRVLRG 8095
QY 208 ----- 207
Db 8096 VPARRDPDAGVDPGTGRQLRQETHGRARAGNDADPGALHRIPIYAQRVERMAEHARP 8155
QY 208 ----- 207
Db 8156 LSCQGRHFLQDPARXADLGLRRRSAGVQLRQGRHRARPRDGEGLPGLQRLVP 8215
QY 208 ----- 207
Db 8216 HQHRAGAGPACGRCPOPPHRAPARPGDGRHLHRHGGPRGLRLRGHPAGGARPD 8275
QY 208 -----TALL----- 212
Db 8276 RVVGRLLRPRRTALLLRDRDRRARPGVCAVGRPAERRPVIPSRLAWKRTXQWRXG 8335
QY 213 ----- 212
Db 8336 NSERSXDSASGTXTRAVERLRCVQGPVVRHLHARLPABPAAPGAGRRVLRGVPGQARP 8395
QY 213 ----- 212
Db 8396 GARRAGRGHPGDVGTARRRDPRPGFRREPAPRRVHRHGARRHORGTPARRRGVDGVR 8455

QY 213 -----RYA----- 215
Db 8456 ARPLRRHALRAPLPRAAALGAARRHLRLRLPLPLHLLGQGRNDRPRYAVHPVRED 8515
QY 216 ----- 215
Db 8516 HPEPOARGLSVDVWLDPRPSLMRALEQAGFTDVEVLHERATAVCDIVAACKAGVPSRX 8575
QY 216 -----ORNCHTSFYLV 226
Db 8576 RCPDEAKGR1DHGCDTGCRGDRAVPTGRLBERGSPGRPGDRPSQGRRTAHHPLHR 8635
QY 227 NAMSRLFRVPK----- 238
Db 8636 GPALRRLLELPRGRXASRPLPRAQGEAGLPGTGPAAPRGTAARRRVRGRPGRAHRL 8695
QY 239 ----- 238
Db 8696 RARPOGAVPGRGAWAPXPRORPAHGLQSEHLPORRTAQRMPPGGSGIAPAPRRRG 8755
QY 239 ----- 238
Db 8756 EGARDTRGSGPQPRRGARRPARPRPGPERQQLAADHHRVRRPRTAGALSAPP 8815
QY 239 -----YINGTKLKNT----- 248
Db 8816 SARQTSNGRSGSLSDSDYVHRLELGGTLHVHGPGLGAARRGTRDPGVPVAVAGTRRRD 8875
QY 249 ----- 248
Db 8876 RPDAGPGARRAGHDGRKPAVLCAGHVHTAAVRSAAAAA PVHRAADA VAOBLRRRRPAR 8935
QY 249 ----- 248
Db 8936 LLEEVDRGAQLXQRGELRRPLAARPGARHHGRGRTGRRAPCRAERLCVARVHRHG 8995
QY 249 -----MRKLKRKQAP----- 258
Db 8996 DRAGAPGLGGPAVVLRVRRRVGRPDVPRGGPLTGRGAAAGRAAAAHALRALQRRP 9055
QY 259 ----- 258
Db 9056 GRGPVAGADPGQTRVHRLGQLGLHLRRRTGAAARHRRGGAAGRGGAAGTGGGA 9115
QY 259 ----- 258
Db 9116 GSAAGRRPAAQLPAGAHPPVLRPAGAPRQQLLHERHGGDSAAVPRQLRHADLRPPY 9175
QY 259 ----- 258
Db 9176 RSRRRHPDPVRPXGHGREGGAAPGAVRPLPPGGGETAGRRRAGTFARRGRRADPPR 9235
QY 259 ----- 258
Db 9236 GRRRAVGERRRGRQHRGSEERVITIGLKQRAVVTGGAGFIGSHLCERLIERLSVT 9295
QY 259 -----VKEQFE----- 264
Db 9296 CVDNLSTGRIANLDALADEERFTLLRADVTEPFSVEGPHVVHLLASPASPLDYLALPLE 9355
QY 265 -----KKA 267
Db 9356 TLRVGSAGTENALRLAVAAGARFVVASTSEIYGDPAEHPQSESYGNVNPVIGRPSVYDEA 9415
QY 268 KK----- 269
Db 9416 KRFTALTAAYARTLGADTGIARLFNSYGPMMRDDGRVVTFFIDQALAGLPLTINGSGA 9475
QY 270 ----- 269
Db 9476 QTRSLCYVEDTVRGLMALMDSSFPGPVNIGATGEMTVRQIAEIAAALAGVELTEFRAPA 9535
QY 270 ----- 269

Db	9536	EDEPGRRCPIETARTKLGWKPEVPLAEBGLKRTLWWESTYPRDMAHDGVNGSSDIRSGA	9595	Db	10616	PRDAAGRRPAQAHPDTARRPGGGQAAQGRLLRVPARRGASARPARARVGAQAAGAA	10675
QY	270	-----	269	QY	306	-----	305
Db	9596	AXQARHPARGCCPGASRNSWMSGGTPPRACRCSRAISPVTSICATVPSGSTSWPRR	9655	Db	10676	RVPGAAGRVGAPGRGALRVPPCGAVPALGRVARDLRPAVPARGVLAVNLLSRLPLGA	10735
QY	270	-----	273	QY	306	-----	305
Db	9656	CPXPRTPICSRSAARASWPASWRPGSAACCACTACSACRWRPAPPVGRGPTSPPTTSGT	9715	Db	10736	LAPAPRPGGEPTAQEVIAERFRERTDPRPDWAYAHLLDLRNALAEQLAGASGRWLDG	10795
QY	274	TT-----	275	QY	306	-STIAVR-----	325
Db	9716	TTWDFRTPPDSRSTSSSTXTPSRSTSTCARSGCCGPAAGSSSTTWTARSGRSS	9775	Db	10796	AGTSPYRDLPCGABLETAEMRGEDLTADHELDGCLGPDGSP--DGLVSTVLEHVT	10854
QY	276	-----	275	QY	326	-----	331
Db	9776	TSTSRATPGRTRSPSRDCPSGAPTWSASSRRRPGXHYPGRTTTRAASASSPCAGNRRL	9835	Db	10855	PDHLREALRLRPGGRLVLSHGVEHGGQDLRWETADGLAAQAERAGFTVDRTVKLT	10914
QY	276	-----	275	QY	332	-----	331
Db	9836	REVPHAGTARPGLLRLPGHLGAERVLGQPDAGRRGGPARVPGRQLLRHRLRARGHG	9895	Db	10915	CGPRGLLLRLRYHGREHGWPGGPVGLLRTTLALADRLRPLRVDYLDLVFAGQGRVEGP	10974
QY	276	-----	275	QY	332	-----	335
Db	9896	RGGTGPRPGRPAHPSLRHRLHDEGRPGAAGRCGPQORRRFPARQPDQPALAGDRLRG	9955	Db	10975	AEFFYLDILLTASKPREPEKPERDAIXPSRPPVPCVCCRGSTCAPPSTVSTAAAVTTR	11034
QY	276	-----	275	QY	336	-----	335
Db	9956	RVPGLARPEGPAGRDGGCAGVGGRARPVRGRLQHRNGRILLRRHPAGRPGAVOH	10015	Db	11035	CITTSCTNPRRNRSSSTPTLSAARTPTPATHRTIPMRSEGXPRRSTAPTCGGPGMRCWA	11094
QY	276	-----	275	QY	336	-----	335
Db	10016	ARPRHREAGAAALRRCDRRHGVVGARAGAGGAPSRCVGPARRLARLLPDLPGRTLQ	10075	Db	11095	SACPRATTAITARTGARRPTGRASTGSPTRWTSPIWPNVPPPTARSRSPSSSTSGSAT	11154
QY	276	-----	275	QY	336	-----	335
Db	10076	AARGRPAEGVRRRTAQCRPAACILGAGPPLGSHFGHRRPASRAPGQRPGRPRRG	10135	Db	11155	PATCGAXCAPTRSAAGVPRPSARNRGRXSTRGTGRTRSCRRRCARTPTSSGGCTSAASP	11214
QY	276	-----	275	QY	336	-----	335
Db	10136	RAARPORTAGRRARAGRQGTARARREVLGRWPRRCPSFSTGTTXPSRSSNGSAPP	10195	Db	11215	SCSPATRCWTCXSRPATRKTSTSTPTTSRCCRRSSSTASRTSRASGSARPATASRSS	11274
QY	276	-----	275	QY	336	-----	335
Db	10196	TPAAWTSVPAARSSRTCAGPRGACTSRSSRCRSTPNSCAATSPRSRYGRPRPPPLAG	10255	Db	11275	SASAPRSGTSSSRTYAVPRPRXTCGRSPRSGXTCGDCGTCSSPARRASTRWPMSCS	11334
QY	276	-----	275	QY	336	-----	335
Db	10256	TASSTWSPATATAACAAVTTAPTREPWRSRWTPYVWTTSSPRTPAWTWSRWTWRAARWA	10315	Db	11335	GPTTCRSPISRTRSDGRPSWPNPSFGATAWTGASPPPICGRXARSXTPLTRPNXSPKE	11394
QY	276	-----	303	QY	336	-----	335
Db	10316	PCMGRECCAGSPWSSNTAATRCATTAPPATTCRCWCSTTSATRCPCRAGWRRSA	10375	Db	11395	EYERVESCPCGCGAAQDLIRDGAGRVPRRARGHVRPARSQRQGDHRRTHPEHPAQ	11454
QY	304	ST-----	305	QY	336	-----	335
Db	10376	STGPLSPPCARSGTSSPIAGPCPPVPPEEKCIHHDYNCPLPLPRGPRMPYGDPRAP	10435	Db	11455	AGRRPCGGRPCDPRRRGAPEHRRHVRVPGGGRGDPDRPREPGDVGPPLPRPGQGAAP	11514
QY	306	-----	305	QY	336	-----	335
Db	10436	GARHRLPAAGCAGRARPAEPRRTGVGAGGGRPDRERLAHPVHARHLDAGAPRARRGPG	10495	Db	11515	YORTAGAVRDRGRQAQAQLLGRYAAQARPGRRIILHRAQALPGRHHRPRPAQFORGL	11574
QY	306	-----	305	QY	336	-----	335
Db	10496	PHRVHGVPGDGPAGGPSAAAARHCDPGRAELHPLHGQVQDHRGRAPAGLAPXRPDV	10555	Db	11575	EDGPPDRAGHRAAHAVPGGRRPARPDLDGDRGPRGRGHPRDPEVPGGRRDRGDP	11634
QY	306	-----	305	QY	336	-----	335
Db	10556	LACRRHRRSRHRRGHLRPRARGQLCLGHPAVOLLLGGRAGHRRRPAQRHVRGDRGR	10615	Db	11635	GGHGPVGRGSGRPLHGDRARGRRGPPYRAGTRPGRRRPRGRDRAAAARAGPRRCVA	11694
QY	306	-----	305	QY	336	-----	344
Db				Db	11695	APPDAGXRVPAQDQCRSKEGAGRMVAVLGRARWAVTDGWWVGRANMINWRRNPGLI FYC	11754

QY 345 ----- 344
Db 11755 MMFPITVVFGVSGMSVAGGDYREFLMPGLFCQAMFVVTISMLAVTSAVSGVTD 11814
QY 345 ----- 344
Db 11815 RFRSMPAQSGVALGRSFADMFSSLIETILLCCGILLVGRWHEGPKALAAALGILLWR 11874
QY 345 ----- 344
Db 11875 YSLIHWGIYLLPPEAAGAAVYVLLPLTMLANTFVSPKPMGLGTIAENPLSSTVS 11934
QY 345 ----- 344
Db 11935 ACRELFGNFGVTGGSWAEHAIELAVWPVLLLVLPVLSVRRYRTLXPKKGNRXXA 11994
QY 345 ----- 344
Db 11995 PNRVPSGTGPKAXSGNSKNAWCRCPCGRSAASTPTTSARRWPPGSARRSPTTSSPA 12054
QY 345 ----- 344
Db 12055 PIAAMATSPREPTRSGCWPSCAARPPAXTGAAARCTPPTSASASMAPTASGRAPPSP 12114
QY 345 ----- 344
Db 12115 PVTPGRSSGAVTAGWFWPSSATVRSARANCWRRSTWRPCGGCRWCSSARTTATRPACRWT 12174
QY 345 -----E 345
Db 12175 GRWPATWRGRPGSAXRRARWTEWTPTPSPRPPRPSNAAAGAAPPSWTAAPTGSTATT 12234
QY 346 PSNR----- 350
||:|
Db 12235 PSRSRWGXTGTTTPRSASGVVTRCAPAPAVTVPAPSTPRPAXSTRPWSSPARACP 12294
QY 351 ----- 350
Db 12295 PTPPRTCTFMPTARPHGPXSDGCTCVHHRAEPGAARDGARRPRLRLRRGRADRPHP 12354
QY 351 ----- 350
Db 12355 GRQPARALRRTRRRHPVVGAGLHQGHRRGDGAATGRGVDPLAALSGVRADRPQA 12414
QY 351 ----- 350
Db 12415 QVLADRWAGGPGHLSGAGLRFALGHRPALRPSLOPVTRRHQDGPABHRLGLRPPA 12474
QY 351 ----- 350
Db 12475 LGGPXPCGGLRAERADGHRGGLRRARPGAARVGPDPHGRGRHRRHRPVACRCPGR 12534
QY 351 ----- 350
Db 12535 RGGHGRGVDRGGSAHDLPGGLDDPGVGEDRTAGRHXRXQPDVVRRRRGHGRGAV 12594
QY 351 ----- 350
Db 12595 RSDGPAPAGDPAGRRGHPARARPGVAAGRGPADRRDPCRPEVTRHDEALTRTGEKMS 12654
QY 351 ----- 350
Db 12655 ERRPRVVTITVGTNELCWLDRCLGSLDSDTTGDLVCYVNDSDAGSVHVRATWPAV 12714
QY 351 ----- 350
Db 12715 TIIRNDRNLGAGANNVGIIRALLESADYYVFLVNPDTWTPDVLRLGJTELAESWPEYGIV 12774
QY 351 -----TAVSEFMKNTH----- 361
||:|:|:|
Db 12775 GPLOQRYDPTDSTALDEFNEWHTHMLGEOHAFAGDGIANHPSPAGLPKGRAPRTLEHAYV 12834

QY 362 ----- 361
Db 12835 QGSALFARTAMLRBIGLLDEVLTYYBETDLCCRARMAGWRVALHLDLGIQHRGGGATV 12894
QY 362 ----- 361
Db 12895 ASTYSRVMRRNRYYYLFTDWDHPAKAARLAGRWLVADLMGRSVVGRVPAATGARETIE 12954
QY 362 ----- 361
Db 12955 ALRWLADRVPWRSRRDRHRLARGGRGAAKASXXGRXERWSTSRCSGRGAMPSPY 13014
QY 362 ----- 361
Db 13015 PHRDTAASSGSSPTXWTGCMNSATRSSCWVPAARPDARGXPSCRRASRRRSSGCGPRT 13074
QY 362 ----- 361
Db 13075 STWSTTTAAVSARPGCRRAPPSARTTSPGRSTPMAVPTVPGRSAPTAGNATTPXSR 13134
QY 362 ----- 361
Db 13135 SRSTRATGPTRTRWPRTSCSSWGSRRTRGRWRPRSRTPAAGAWCWPGPSPGSRSTST 13194
QY 362 ----- 361
Db 13195 RSLAGTARPSRSVRWAAAGGSCSPPHTRCWPCCRPRSPGRGAASGASRVRRWSPPPPXA 13254
QY 362 ----- 361
Db 13255 ALPSWARATAWPRSCPRSAKSWDTAPTSTRPTKHAGPMQCEHPTRSGVRRSGCGMXXRS 13314
QY 362 ----- 361
Db 13315 PNGMWSSTAGCWPERPSEGPRPGAGGPPGRHVVSPPCARGAARGRRPLRPPXPRAHR 13374
QY 362 ----- 361
Db 13375 PGASGRXDWGPNRAERAGRAPRPHPRRGAPRRPGPPRRSPAPGYGRRRCAGAAARDR 13434
QY 362 ----- 361
Db 13435 GARRAHSAGRCPRWPRPAGSICXARAGASAGSPRPGARALPRESSAASBWAHRSA 13494
QY 362 ----- 361
Db 13495 RPPGCPAAARRSPGTAGNAHSRPARGRISPGARSAVAGPACRNPGRRCRSPXXSRA 13554
QY 362 ----- 361
Db 13555 RRTGDRRHRATPAPCRPRAAAMRIRRPGRUGAPRRSQFPFPYGGSFQAFVNDTF 13614
QY 362 ----- 361
Db 13615 LFGIWHRLXPENSELTRLPLVAVNPWPSGAPIDARVILLSARRMXTGAGCYDWTQDCV 13674
QY 362 ----- 361
Db 13675 GCRRHLLRRAPRTGLASDLRPCGHRXSECYPPRFHRGHRAFLSGRENAGGATGHS LFG 13734
QY 362 ----- 361
Db 13735 SGRRELPGSTGLAGTGPCDSWFDGCGHCPRPNNADGGRESAARVRRTRTGMGADRRPG 13794
QY 362 ----- 361
Db 13795 DGGPRLLRGRQARSPAAGHRPGPARPAPRAAGGQPPRRRPRGTRPPGLPGGVRTAGG 13854
QY 362 ----- 361
Db 13855 RLFOATGRPSGPHGLPRPARGFYRASGGRRAGRLFRPGRVAPGAPCAVTADLLPG 13914
QY 362 ----- 361

Db	13915	DRTAVIGATGFIGSRLTARLTADGPRAPAFPPVVGRAAPGLAEADIVYFLAAGL	13974	Db	14995	CMSRMARRAPTWRPGRWDGTGRSRRPRGRPTPICATPSWCTRRWPTPAPARASWFSRTC	15054
QY	362	-----	361	QY	376	-----	375
Db	13975	SPVLAERRPDLVEABERLLVEVLEALS RAGHRPVFVLASSGGAVAYAPEVPPYRETSPTR	14034	Db	15055	CPWSSNSTGPTGGTPRSSRQCAGAWARTYRDRERAGDEMAEKQAVAPPMTDLIRLVFGGM	15114
QY	362	-----	361	QY	376	-----	375
Db	14035	POSVYGHAKLERELCHRSADVAVVAVARLSNVYGPQRALRGFVLPWHLTAAARGEPV	14094	Db	15115	ATQVVGLAVRLPLDAIGEGERTADGLAADFESEPAAMNRLRLGLAALGVLRSEKPGVFA	15174
QY	362	-----	361	QY	376	-----	379
Db	14095	RVFGDPHVVDYVHVDVTRFLALGRVGGGRLPAVVNVGSGVFTSLGELLEIVSGVTG	14154	Db	15175	LTPVGELELRADSPSFHSLARMLTDPAVATAWQHLDHSVRTGPGAPDHFVGRDFFAYLAD	15234
QY	362	-----	361	QY	380	-----	387
Db	14155	GSVAVMERGRSFDROGNWLDVABARAEALDWRAAIPLTEGVRCQWVRLHDAGRAPSAST	14214	Db	15235	DPDLSWLYNAAMSQGTGGIAGLVAHQDFSGVRTVVDVGGDGTLLAAVLRAHPSLRGVL	15294
QY	362	-----	361	QY	388	-----	387
Db	14215	RXTVRRVRSDASADGAMPGSTLVPTWXSASTLXSGPGPADSQRRAGHGVETRGG	14274	Db	15295	YDSAAQVAGVAGVLAAGVADRAVEAGDFFAAVPEGGLYLLKSVVHGWEDERAARILAH	15354
QY	362	-----	361	QY	388	-----	387
Db	14275	VVEVGVHLAGSLADRCGGHPGRRHQTGVLGRREQADGVLRQGVGLGVRLIGAVLVRA	14334	Db	15355	CRRALPAHGRIVWVEHLLPDTVPADAVPTTYLNDLMLLVNGNGLERTRGDFEOLCAAAGL	15414
QY	362	-----	361	QY	388	-----	387
Db	14335	GDGVVDVGOSRQDFGEGHRAVDLLVLGRGALRGHRVQGLEVGGLAGLLLVHAGHQEQPL	14394	Db	15415	TVBEVTLPLAGTDLMLIEAVPASADPLGXAALPGISPCSGASLRDRAEMSHHPARDASEFT	15474
QY	362	-----	361	QY	388	-----	387
Db	14395	VGVAVERLVERPGLLLRDVLVHRGHRGQAAVERQVADGQQRVEEAGGVAGEVWAG	14454	Db	15475	ILLARTVIYLGSSRTVSGDWRYALFTAGRXFCLIGERSLRESLARMALTSRVNSNPMKS	15534
QY	362	-----	368	QY	388	-----	387
Db	14455	AVEAVGDVGEVLLHPYARLVDAVRHEVAHGRGFLDQGOHRLQGVFPALQEPVLVADE	14514	Db	15535	SEWCAGRAGMPATVGPCTRTTSEWXAATCWTTRAXSLLSSSPAPPSTTSAPCRSSRSPG	15594
QY	368	-----	368	QY	388	-----	387
Db	14515	TDVAVPAQRDPVPAVRGADDVADPCGRTGQORALPVAPDRDVAEDLVRAAVLQVLGD	14574	Db	15595	SSPPLPRXPVAVCCGWPRPSASRGSVPPPPAAASRASIRPPPRCSCSGSPRRARP	15654
QY	369	-----	368	QY	388	-----	387
Db	14575	DAGPARGVDDVVELLADQLTGVGQPLRPSRAVAVEQHALDRDALGDVGCALLDRVLQEHV	14634	Db	15655	RPSTRAATGWPRRSGTTWCRTTASRTGCSAGSGCSTTSHRSAGRSSRARCRCGWPPX	15714
QY	369	-----	368	QY	388	-----	387
Db	14635	VEVAVDLPFADVGTEBDRPVDPRVLLAPAQHAVLDRTAGVDLLQRADEVDCGAARGG	14694	Db	15715	SWAXWTSSGWATWRRNATGASQGTTCAXRXKCSPTSRRTSSSAPGSGTCALWISPSR	15774
QY	369	-----	368	QY	388	-----	387
Db	14695	DRLADVAREGLPFOXLHGHALEOGGRRGRSRRPTADHEHFFGHGAPHXXPRFHVAHIC	14754	Db	15775	TSQWTGATTWRTVPGCCARRRWNCSAPTXYAALGRSAGSPRSRNRNWSPXWSTATCACSP	15834
QY	369	-----	368	QY	388	-----	387
Db	14755	GKXKTSGTGNCCELLRVXSSRSSGISSPSSCATMXHPLRDXCRAPRPSCKGAXWEPAYXP	14814	Db	15835	TVRTPPAPFNSRNVCGEPGPRRRRTSTSPPLGGVPGVRRRCVHRLVRPVLRRVAV	15894
QY	369	-----	368	QY	388	-----	387
Db	14815	RRTXELRLRLHPLAGCVPRGARRRGVRPAVATGHPRRPGHLDAGSGPAGGARRRRVP	14874	Db	15895	PGDAGAADRAAAMAVRRRHGKPRVRSLLRRRPGAGAHQLRAGPRVEPAHHQHPAGDRHP	15954
QY	369	-----	375	QY	388	-----	387
Db	14875	ARRQHPDAARRLRPARRGGFVAARXRRSARSGFFXPGRPRRTTACTSTPASWPRTR	14934	Db	15955	RHAGOGGRRRSALPRAGQDRARQPHPAVADRPGHPQQLHPHPGQPHPLSGVLHQGR	16014
QY	376	-----	375	QY	388	-----	387
Db	14935	TPGTGVPWXTXSRRTTTTRSSATTCGPAAGCCSCSRTPARITPRXSAAPTWCRR	14994	Db	16015	GPHGGRPAVRAGRVVSAQAQPHGCRPGXGRARQGLPLAAADRATAAAAHRRPQGHG	16074
QY	376	-----	375	QY	388	-----	387
				Db	16075	PDRAGLPPGGYVPRILPAGAGRXRRGGTAVVDGPGHLPERHGGRAALVHRGOEPAPTDGT	16134

QY	388	-----	387	QY	397	-----	396
Db	16135	PDAPAGPARVARHSGADHPGRGAALQHHRGDRPGRXPRGRMGPAAPVAGARSRRVDO	16194	Db	17215	SVTSIRCATTERRRSGSGPGRRASTTEATGRAKEVGGRRIRPLPLPWRKSRASGKXGSAS	17274
QY	388	-----	387	QY	397	-----	400
Db	16195	GHRGAPARPRVPAPRAGRHGDPDRAVRPGELSAARAAVRLRPDRFPGLQCRP	16254	Db	17275	TPRRNRTRSNRSGSGSARTPIITRPARSRGRSKCCATATIRNCGSRPGSNRTVEPAD	17334
QY	388	-----	387	QY	401	-----	400
Db	16255	GRGGRALLPLAEQLSPGRRRRLPRSRPXLLIGDRPPAHAPAAARLLRQRRGPEPARLP	16314	Db	17335	LGLQPVGLAEGPPRADVERAGVRSNADPQGRASAPAVDATPARVACGNVARGAGARW	17394
QY	388	-----	387	QY	401	-----	400
Db	16315	AELVTPXTSVPCGWAFSAAPSPDAVPCRPXRCRXXPPSPPCARTTPVRSRPGSAA	16374	Db	17395	GGVSGRRWPGAGGFTRRRRRWSITITHAPSRPYGRPNRSMGNWPGASWRSRWPGCRTPS	17454
QY	388	-----	387	QY	401	-----	400
Db	16375	RPSSATRRCWARMSTRSTSPCPPDCTTTGSCAPCARESTCSRSRXPPATRTPXRRSAS	16434	Db	17455	RAGCAWRRRAGSRGRSARCSPTPARTGCASSAGSGSRPTTRWPPPSGSPRLYGTASPYT	17514
QY	388	-----	387	QY	401	-----	400
Db	16435	RTPAAXXWRTXPSPTCTTPCARCWTAAASVNCAPSPASSGRHCLPATSGTSRSWPA	16494	Db	17515	STWTGSPGPPHWSRSCCTCWRRAPRHAGPSRRAGXRRSRARCWPRRRRQAGCACCSR	17574
QY	388	-----	387	QY	401	-----	400
Db	16495	AACWTRACIRSPRACSSGPTSMWSGRCGWEPKXSTWRATCCVPTAARHTCPSASS	16554	Db	17575	GGPHGDGRWRPWRSAHXPSRPTTTVCHRCSPRPRSTCCAAARAPRWPWMTSSCAPPSTT	17634
QY	388	-----	387	QY	401	-----	406
Db	16555	TPTAARTCGGARAGSSSTVSPRRPTSPAPFGWNAPTACASACRPTTSSPAPSASSG	16614	Db	17635	ACSPSSRAAPTPPRPRGSCRTRWPSGSGPRSTTRTAGSHCAATRPAPASPSBSAGRS	17694
QY	388	-----	387	QY	407	-----	406
Db	16615	PSPAAGRRTPAPMSHGPDWVCRAHGCSATAGOEQADRRERVLVAGTGFGVRQL	16674	Db	17695	SATRWGASRRSRPSHSRTXPEGRGRATSWWSVRPACXSTGCRPRSAVCCAXHRCPTPTG	17754
QY	388	-----	387	QY	407	-----	406
Db	16675	CADLTAAGAEVAIARRVPDFPLPGRLLVDVTASPGELADVDSVRPHTVVAIGSNW	16734	Db	17755	RTRSPTRGGAASRPPSTTRCGASRRPGAAARWACSSWTRPTTYRIPGPAAPWPSPSGRST	17814
QY	388	-----	387	QY	407	-----	406
Db	16735	GIAERDMETNCAVTRRLLDALRRATACRPYVVHLGSLVLEYGPTPPGPEPTRTASPPRTTA	16794	Db	17815	ATACCSXPVRPWRPTASRASAACWCGSSSPSWRRPSTTTAWRAPRSARRSHRSCAATSR	17874
QY	388	-----	387	QY	407	-----	406
Db	16795	YGKAKLAASQAVLEAAAEVGEAGVLRIANVAGPGTPANVSLGRVAGRLAEAVTRDWLPA	16854	Db	17875	TSSPNSRRSSTRTSGRSPARRTRPTARPCADATSWRCAGRETRVRSRGTCTGCARSLS	17934
QY	388	-----	396	QY	407	-----	406
Db	16855	VVELSQLRAHRDYVDVRDSDAVLAATRAIRPGLVVPIGRGEAVPVRWLVDLLVEVSGVP	16914	Db	17935	RPPTGTXRSMWCSPRSGTYWQWYARRSRTRPGQRKQARTSRGPKRGEQPMRGARRWKG	17994
QY	397	-----	396	QY	407	-----	406
Db	16915	AEVRELPAATGAAGDDWIQVDPGARELLGWTAVRSLHESVTGLWADTLRROAIPQHA	16974	Db	17995	QPRTERARRRPPYCTHERRRSGRGRACDRDGPVCVSGSRSAACRPPADSSSTSP	18054
QY	397	-----	396	QY	407	-----	406
Db	16975	PRPSXALSSDRLPRRPPSGRTTGIRROGPAGGAGSQTPETGPHHSTRPAPQVTAARR	17034	Db	18055	RPVTPYCSRRSRQAEWASTCRPPPWKSSASRSSRSPNTRRSFVPTAWARSARSACTAS	18114
QY	397	-----	396	QY	407	-----	406
Db	17035	HASAYARTSGRLPSSRRRANRIGESFRTRGVVTPPRLISGALPVATROMPSPATRGFS	17094	Db	18115	ARKGWTNACGWCWRTSPACSTLTTPAAARSFRRHTPSTSRTPWPATSWRNRHGWGRF	18174
QY	397	-----	396	QY	407	-----	406
Db	17095	AYRGRCALRPGASRRTAERSSSWRACPRGSRTRSPTSASRCXSSSRVPSAPCRSRDGP	17154	Db	18175	RSPLRFGGGGEGPPGGADLEGGAVFGRVGLVADVGVGAEGNLDALAAAVARLEP	18234
QY	397	-----	396	QY	407	-----	406
Db	17155	GPARTAACTRRXGAPAASSPAATAPSPATRRRRRXXSWPRRRTRPRCGCRWATTGSLPC	17214	Db	18235	LKCGAHLSPTRRSRLFDSCGVMACDRSIAYLRARMDTSSGPETHWLPRWLSVXARCWXS	18294
QY	397	-----	396	QY	407	-----	406

Db 18295 GVSRRVKGPSACAEVRMGRNWRASFGSSAHRRWRSMRTFGPPSSMYTGSHTMKL 18354
QY 407 -----MGFQ----- 410
Db 18355 IIGGHLRCRISWRMRPSVLASSSSSGYGGTRSRRTARAXSRVWSRPGRTALSYQHREMA 18414
QY 411 ----- 410
Db 18415 SRCTMYCSANGSNWISGMFSTPAANVPFVSRSSSAWASSSRGRLPCSIDAMVSSA 18474
QY 411 -----RTFIDPL----- 417
Db 18475 RTALPRELCVKPALRAACICLPTIRWDRFFRPGSFLEFMLAXLPTRTHKDPVSTRKS 18534
QY 418 -----WD 419
Db 18535 EYCTHXKPIITLDIRPVNDANELPHQDOFYARNRRSIPAAARNLTSWSLKYWGSLDWD 18594
QY 420 YLDSLL-----FLDEIR----- 431
Db 18595 RADDVLLCVSELATNALLHGVPGHGFLLRVYDGDVLRVEVHDSGTGVPRTDDPDEGG 18654
QY 432 ----- 431
Db 18655 RGLLVGALADKWVGERTPGKVWAEFAVQGCAIPVEASIFRLTHRRGTVSXHAPHS 18714
QY 432 ----- 431
Db 18715 AASPQOAHRLRTPQMAVWRHTRRRLAALSVTATAALAASLVSAALPATAAPBGRIOYAG 18774
QY 432 ----- 431
Db 18775 AADAVADSVTLNKKADHARSGSEAGRALVEKYGADIERTYKKALNGVATEASEAEAKALA 18834
QY 432 ----- 431
Db 18835 ADPAVASVQNRMFSDATQTNPPSWGLDRVDQKNLPLNSSYTPDSAGQGVYAYVIDTG 18894
QY 432 ----- 431
Db 18895 VRITHSDFGGRASYGYDAIDNDNTAQDGHGHGTHVAGTVAGNAYGVAKKAKIVGVRLNN 18954
QY 432 ----- 431
Db 18955 SGQGTAAQVAGIDWVARNAVKPAVANMSIGGADTALDTAVENAMASGVTPAVAAGNES 19014
QY 432 -NPSLRSP----- 438
Db 19015 TNASTRSPARVTEAITVGATTSSDAKAGYSNYGSLDLFAPGSSITSAMNSSDSATNTIS 19074
QY 439 ----- 438
Db 19075 GTSMATPHVAGAAALHLAANPSATPSQVATALTSAATTGVVTPNGTGSNRLLYGVRHD 19134
QY 439 -----TYVN----- 442
Db 19135 HPPGPRFENTGDTYISDNSTVESPTVTVSGVGNAPSALAVEVHIVHTYIGDLQVLQIAPD 19194
QY 443 ----- 442
Db 19195 GTAYTLKSYGTGSSDNINTTYSNVASSEANGTWKLRVSDNANYDTGRIDAWALQFXPS 19254
QY 443 ----- 442
Db 19255 LTSPSRPGTAPGPDPANREDDPRPARRLROGSRGPIGGVGTGCLPPSPSRGCPQLR 19314
QY 443 ----- 442
Db 19315 MLRRWGAYHPGPAVAYGKGTCTGVSARSDBRQHAAPGPRVPPVGRTPPGAATHNAT 19374
QY 443 ----- 442

Db 19375 QGYAHPHPPTGRRRHAAHAPHHRGITTGVPISSLAPRLPFAQSYASSRRPALPPEG 19434
QY 443 -----LTPPE-----HRR----- 450
Db 19435 ARYPCYLAPPDPSNRKGPSSPKGTRRASTRHRRHTAHPGPRDRTTTTPTAPTAP 19494
QY 451 ----- 450
Db 19495 TAAPHPSTASPSAPWSSASSAFCRPXGWCWDXSRWRRSRGASEARAWRXPVRCSPASDS 19554
QY 451 ----- 450
Db 19555 SCGRCCPRASPTSTASRTTAAARAPPTSPRATASRTTPWARPPTTWTRCPAPPR 19614
QY 451 ----- 450
Db 19615 TTARCSRSSPSAAPSRATTPSPPTTSVTRSPPLRHGHLGRTGRRGLLLHPDPELA 19674
QY 451 ----- 450
Db 19675 LRRPRDLFVRRHAGRRQADRLTAQRDDTTRRRPGRLPLRDQRRHRAVRGAGGIRGRP 19734
QY 451 ----- 450
Db 19735 GREQGLGGRPRRRARRADRGLRGHSWPKGAAPVADLVEEMEDARKDWAKAAAATNADTY 19794
QY 451 -----AVNLSTS----- 457
Db 19795 YTYDSGYEYVDGPATVTARKALGLDTPVPSYEEDPGTSGSGSGGGGLNVXAEVXAR 19854
QY 458 -----NSLW 461
Db 19855 LXRGESACVKPSARTSHHIEFSGLCGLHGTTCQCAVAVCTSXW 19898

RESULT 12
ABP76681
ID ABP76681 standard; protein; 19938 AA.
XX
AC ABP76681;
XX
DT 26-FEB-2003 (first entry)
XX
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 5.
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
XX
OS Streptomyces viridochromogenes.
XX
PN WO200268436-A1.
XX
PD 06-SEP-2002.
XX
PF 24-AUG-2001; 2001WO-BP009815.
XX
PR 25-FEB-2001; 2001DE-01009166.
XX
PA (COMB-) COMBINATURE BIOPHARM AG.
XX
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX
DR WPI; 2003-018650/01.
XX
DR N-PSDB; AB237516.
XX
PT New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes.
XX
PS Example 1; Page 68-301; 319pp; German.
XX
CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human

CC or veterinary medicine, particularly where caused by *Staphylococcus aureus*. (I) are more hydrophilic than known avilamycins. The present CC sequence is that of an avilamycin synthesis enzyme from the *Streptomyces viridochromogenes* Avilamycin A biosynthetic gene cluster (AB237515-AB237516)

XX Sequence 19938 AA;
SQ Query Match 21.9%; Score 536; DB 6; Length 19938;
Best Local Similarity 1.2%; Pred. No. 0.011;
Matches 218; Conservative 96; Mismatches 135; Indels 17947; Gaps 66;
QY 1 MGRKE----- 5
Db 4 LGREAGVRWQXTVDLLAEITRDAQVHRLLTESHVPLPSACTDSNSLATLGCVPQAKARE 63
QY 6 -----MMVR----- 9
Db 64 SLDVNTCAGGLYAGTTSPLXPRSHLSHIEAAAAALAPAVAGAGVLLVRRHGGVQAQS 123
QY 10 -----DVPKMFVLISIFLLVSFINCKVM----- 33
Db 124 LARGDGRGADV-----LVSAVVVRVVRVRRRLRPVLPGLVLLDQIGD 171
QY 34 -----SKALY----- 38
Db 172 RLGGALRPVSAQPSICSPSTAMTSPAALFAARSSAYSSSYSASTALVAERKATW 231
QY 39 ----- 38
Db 232 SASVVGSLRSEPVSLPSSCVAPNQRVISRPKQLWRVGMKXXTSASSGTAQVSMAXRR 291
QY 39 -----NRP 41
Db 292 RARNTCPRSGXVRRRGGXRPRTPRRARRSRAPRRRPPPCPSRSSRP 351
QY 42 WRG----- 44
Db 352 WRGRGCPKRRRXPXSRRRRRXSPWTATPSTGKRVRRRAPNRLSPCCLARPASXAPA 411
QY 45 ----- 44
Db 412 RSVAPALRPAESRGCGPGRWRRRRVRRGGRGCGSRGGTGTVPGRVRRAGMATV 471
QY 45 ----- 44
Db 472 AGTGPARSALGAAPGTLLAAVGVGAEVWTRVPCSWFCRQRAAGTGVRLGENXAGGAGR 531
QY 45 ----- 44
Db 532 NGDPEGDPVAVGVVRMPAACCAGWGVVRVALGCVVGRAGGCPSPRRHPWARRCV 591
QY 45 ----- 44
Db 592 LTPVTAGHPTRPLPAVRDCGPGVRAFSQAHXIGTTPARRANGEAPRTHPAQRASTP 651
QY 45 ----- 44
Db 652 LPEAACRGTWVLA VGRVTRCRAARTTRTGGGLELQRPVYAPGVVVRVVAHPQPR 711
QY 45 -----DQLKLEI----- 62
Db 712 AVGLRGGVHRVGVVDVVRTAAGVRLERVRVAVGGDQDLEAVDVRVDDVDLYEGRGR 771
QY 63 ----- 62
Db 772 VAGDAGRHRGLDRGVADVVTGVLEARARRVVPADVQBPVGRARAGVDDAGGGG 831
QY 63 ----- 62
Db 832 RORRGDLGRGCGVGEVERGGARHVRGRHRTGDBGVRRVAXVPRGDRGTGGEVBDG 891
QY 63 ----- 62

Db 892 AVVGVAGLGVRRARRDRDGLDAGCGXPRGGVGRVLVAGRHEGDAAGHGVA YGRVEGRVG 951
QY 63 ----- 62
Db 952 AAAEGHVGDGRLDGVPDGDVADGDDLGGGALAGVVQHADAYDLGLLGDVAGVADGRAGHV 1011
QY 63 ----- 62
Db 1012 RAVAVAVLGVVVVVDGVAVGGPAEVAVGADAGVDDVRRHSLPGRVVRGVRAVQGEVL 1071
QY 63 ----- 62
Db 1072 LVHAVQAPGGRVCLGRVDAEHAHLDHGGDRRVGGERLGLRLGLGLDRVAVOGLLVRLDVG 1131
QY 63 -----LROLETTI----- 70
Db 1132 AVLLDEGASGLGTRTCVVRLEVDVAVGHGVCRARVLDLTALGGRGSGESDEARGEGRG 1191
QY 71 ----- 70
Db 1192 RHABGQSSPGMTHHCHLRGPPQSVVRLWGRRGVGRGMSTHCASPVROPKDXGLHRNCT 1251
QY 71 ----- 70
Db 1252 RALHGEGLPDDFAGRS LAHAPLVGQADQEQTAAPFVGVVGGSGYARS AVVDLDAEDVAV 1311
QY 71 -----STKYNVSKQPV----- 81
Db 1312 IADAKETMAGRDVAEEGVGRQLAHAEYVVGVPVGEAPVLEXPGGQVAGSGDGSAVAC 1371
QY 82 -----KNLT----- 85
Db 1372 VELVALVGFVVCVVGTVNVAQRDWLLSVRAPVLGCTGHRVLMGACGELGQHELEAAA 1431
QY 86 -----MNTFFQYXILAGP----- 99
Db 1432 EBEPPVDHADGQADAGRQGGGLHAEPFGQ---SGPRRGDHRVDRTGEAPAAATRPCA 1488
QY 100 ----- 99
Db 1489 GRTPRHEGHVCGRGEHAGDRPVPAVGAHVHARAGAGHLVPLVRQRRPARPAPDRLRPR 1548
QY 100 ----- 99
Db 1549 GPARTAVRGRAGSQDGGSHRTFGDPAQVPADDELHRVGA VHRTRGAESP RPAP 1608
QY 100 ----- 99
Db 1609 PPTRVRRTPGARPPVSTHAHLRTRRTGRPLHPPGDARSAPQLHREPTWQPVGFWPGRGV 1668
QY 100 -----IQNYSITY----- 107
Db 1669 HPGAQVCDAAVTGHDHDSAGVEEPARPSGGRXMXSTALEWFKSSYSGSEGGQCVEVALCPHTI 1728
QY 108 ----- 107
Db 1729 HIRDSKNTPEDGPTLQVSPTAWRAFTSATTEXARRPELTPTPTVPVPPRCGPGRPR 1788
QY 108 ----- 107
Db 1789 RRPVSTRPRCGGRKRRTGTCSPASAPVRRPPLPCRAGGARGSPGPGCGHGRPPGVR 1848
QY 108 ----- 107
Db 1849 RSAXAAARRGSPRRRPACKGFLPLRSARAVRRDRARRRRRRAA VGRRAARCPRSARTH 1908
QY 108 -----LWDF-----YSTQLRKPK----- 122
Db 1909 ARPAPAPCAAPSAPSAPVCPVRRRSPGPACPAWLFPFPPGPAHPLWLAPLRTPARPAG 1968
QY 123 ----- 122

Db 1969 LLPLARPRXAPRVLPVPRGTRTPRSARSRPPQBSRAAGAPSRTPADARTPPCAS 2028
QY 123 ----- 122
Db 2029 PXSCRRARPRGRPPPCWAPPTRPCAGAPGVXGRPCAGCADDRPVPRPCGTPWSPRRR 2088
QY 123 ----- -YVY----- 125
Db 2089 ARRWPPPARAGGAPCGTTPRRGSPWAYRSGTGARGSAPTTRRRPWSGGSGDSRSGPRP 2148
QY 126 ----- :----- SOYNHAK----- 133
Db 2149 RRACPPRRPAPGSPAARRWXPRRPPAPSSRTRPAGRGRAAVSRAAHCAHSRSPSSXS 2208
QY 134 ----- :----- 133
Db 2209 ARWPDPPXGSAPRLLGARVRWPGSSCRAPSSRRSGSGAGRGSGCRRPGSRAVRPG 2268
QY 134 ----- 133
Db 2269 ACRAFCSXPAPRPPAGTTPRPRGRGRPARRAGRSTRRSAAASRPGRRPSGGAAG 2328
QY 134 ----- 133
Db 2329 RPRGRAAVAHRRGRPARSVRGSPRRPPAPVVRPAPRTACAAGLPPAPPRPAGRASSS 2388
QY 134 ----- 133
Db 2389 APRPPRRPGVPRGPPPARAARRPVWSGARACPRCVRCCPVQPRRSAGRRPPSGRRS 2448
QY 134 ----- 133
Db 2449 AARPRGAGAAGTSRRRAPGRSGTRFSPPPPGAACPXGPTAGPPGXPARAGFVAHGSV 2508
QY 134 ----- 133
Db 2509 PGPPVPRRRVRDRPPPPAPPREAARPPPPPAHSAPSRPRSHRVPASCHPGGRI DGW 2568
QY 134 ----- -TTF----- 137
Db 2569 CAGCPATLTIGRRANTCPLVVGSGALGQADRLKAEIGRLNGFVRSWPAPAVDTRRSRA 2628
QY 138 ----- :----- 137
Db 2629 ASXPTTGPRAGRDRGPGRGAXRSVPXSRPPVPAWRRCRALLPRRSALPPWKRGXGADSA 2688
QY 138 ----- 137
Db 2689 PPNLFCAAGLSRRSPSPWSASPTSCRCAPDRGRAROXPRRRPTAAAAGSRPPPARPG 2748
QY 138 ----- 137
Db 2749 SAPPSPGRAGSRRRRRRSGSGSPGTGRCPCWPRPVPSRTARCENPARSAPXCRS 2808
QY 138 ----- 137
Db 2809 RRPRPRAPRTGPGRTSFRTCTPRSPWSQGSASVREAPGRARHLPRGDGQRPADQPR 2868
QY 138 ----- 137
Db 2869 RGDDPSGTRALABFVRPGVAPGGRQASAGAVRGVPPGGGDLRGTTGCGVVRSSLGRL 2928
QY 138 ----- RDP----- 140
Db 2929 ALRAARWALSSNRRPPGGRAARQTGIXROGLGRPLRVLGDCLPAQVRVQPGCHRLVQR 2988
QY 141 ----- :----- 140
Db 2989 ADGCPAEFAGPGVDLDVPPVGRARGGGKLPDLGGAAHLDQIDEPHRRHGLAASDG 3048
QY 141 ----- 140
Db 3049 DDQPRNTGAGGGQHRVRDVAHVVDVAVGAQLGQLHRRHVRHVGDLGEPFGHPAEQADR 3108

QY 141 ----- 140
Db 3109 GARPGHVGDPQHARLHDLALRGGLQHGLARRBEGLAVGRRRPPRRRPGRFTGRGRAVLOH 3168
QY 141 ----- 140
Db 3169 XAEVDDVGRTARGPPORVQQAAGRHRRAVGLHVFPDPPVAADGVDRHRVADGVHDVHVGQFPG 3228
QY 141 ----- 140
Db 3229 RRGDGDQDEEPARQCEVRYPPGDRHLHSGRGQVRAQLTAHEPRPTSHENALALIGHLLL 3288
QY 141 ----- 140
Db 3289 LTRCRAAVPGPAPDLPVGVXHRRCAPPTRRRXRPGHSTGRRTGRRRPAGRARAPRR 3348
QY 141 ----- PCGR----- 144
Db 3349 RVPAEPWGGSRAGAXRHGRXSGPRSPACTRSGRRARSRRTRVPCGRRRAHTAARRPPRR 3408
QY 145 ----- 144
Db 3409 LSPRPSPSAASPRPHRGARGTCSPPREADTRPRPAGCRRPAPAGTGRCREAVAGTRTRMXG 3468
QY 145 ----- 144
Db 3469 SAVHRCRPRPASRRARRAAAGXRGSPSASGSGSPRGPPSPCPCGSGRSAASPRARAPP 3528
QY 145 ----- 144
Db 3529 RAGRARSRGAVRWAGRRRPRRHGSPAAPRRRRPRRTARRPHRCRCPRSGRRRSPSA 3588
QY 145 ----- 144
Db 3589 PAPSARPAGYGVGRCRCSRECPAARDARSSWGHLQCRQASRLRASTLTXXPCRRSRVSNW 3648
QY 145 ----- 144
Db 3649 AVTQXSGTATGKSSASTREXLFCEWXKRPPSSARTALXRRRSGEAVRTXSRNRGGPC 3708
QY 145 ----- -VPSMT----- 149
Db 3709 GXPSGTHCTVGPISMPSSRPGNRAWTSCATPMTLTINATTPRCGNSSGWSHATS 3768
QY 150 ----- 149
Db 3769 SPTWAVTPMMLKCRPSSWVICSGVPRHPCRSRRGVRCVSWCRLLASVNQSTAPMSPRE 3828
QY 150 ----- 149
Db 3829 VFWPVDSDSTAAPTSTGTGREAGNVAGROASTVRVSMLTRPSMRSSRRSSVSGSQWXP 3888
QY 150 ----- 149
Db 3889 LAAGTSATWTTMLRLRLRNRHAPCSDCTSTRVPAPALVKNRXXRVLRPWWMRVXLLR 3948
QY 150 ----- 149
Db 3949 VAVTVGDSWMLWLPGLSILACTRQCCTPSTSLTSMRMPISGLWLMGHLHSGTGXPXVRT 4008
QY 150 ----- 149
Db 4009 RRPSTAKKAPDSCARFPVSASNRHRRSSVSGARVTRHCAAAEHRANQSMNSAPATSDRYA 4068
QY 150 ----- 149
Db 4069 SEGRAPARARPTWPAGRLTTDRPVRRPRRRPHCRXAGACRPSWRPVPRTRPASS 4128
QY 150 ----- 149
Db 4129 RASAEGRSRRRAVPPPPGAAAGSPPPRGRASPARRRRPPDRAPGXSRCRGRGPPA 4188

QY 150 ----- 149
Db 4189 RPRASSCRARSLRSPSRVPPPGCPAGACPSGASRRPPRATPCAXRTPGSGVTRGX 4248
QY 150 ----- 149
Db 4249 TGSXPRTSRTTRTSSAPCSARTPWPFRRCAGAGRAWPSAGTPAPFTSPRWRPDSAG 4308
QY 150 ----- 149
Db 4309 RRCWNSGNCXAPXCGWPPASAAHPCSPRQRWAASRRPGRTTGRRGCRSRWGRTARQQ 4368
QY 150 ----- 149
Db 4369 RSGSGPASRRSPLGCPGAGADRRTGTAGPARAPFRRLHIRKXVDARGKCHPGQFPXX 4428
QY 150 -----CLS----- 152
|:|
Db 4429 ALAXTESSPSREXCVSPVTGNGTGRAQVNHCSRKQYSEFTSPCWMGRPPGPIPOXGSRA 4488
QY 153 ----- 152
Db 4489 RXNSRGRGAERVSRSGHGLDQOVGSGQRDRDRDQARRRAELFEVATRALQPVADQQ 4548
QY 153 -EMLNYSKRN----- 161
|:|:|:|
Db 4549 VEVVEGRRHGVRHVGQQVLDHDDPAVRRQCPAAVRQNPGRPLVLPAMDDALEQVEVT 4608
QY 162 ----- 161
Db 4609 ALGYGEEVAGLHRAVDARRQHLFCLRDPGRRVVQHAAQGVRAQHRGQORVPAPDV 4668
QY 162 ----- 161
Db 4669 HHGADAGEVLVRDQADPTGALAHRRVVEPRQVGVVRQVREVPVPAEDVVEGRAAGADAV 4728
QY 162 ----- 161
Db 4729 VQVLPGRDRGIGEPQXVEGRRVAVGATQPDRCQEHAGLGLAQHSERGQAAQPVHG 4788
QY 162 ----- 161
Db 4789 SGFALEVGGQVGGPFALPDRVGPQPHRQAHDLGGHSAEHQDPQVRPGRHGLLRHLV 4848
QY 162 ----- 161
Db 4849 TGELVAVCPGPGALPARPGCTARRAGRVRAAPRAAGSAGXPXAGGCRWPPACAPGRG 4908
QY 162 ----- 161
Db 4909 GTDRRPRPSFWPTXAAAGVPPPPRPPGRCPPRHTAPAAAARPGGSRRGSAWARPPRCRPT 4968
QY 162 ----- 161
Db 4969 GAGAAAPGRGTRTCGRSCRSPAGSRPAHRTSPGRRPRPRPGSWRRRCASRSPRAASS 5028
QY 162 ----- 161
Db 5029 RETAPGTCRPASAGCHRPSPASWSXAACSVGLAARNTSSRTPSRRTSRVQVAGSSGS 5088
QY 162 ----- 161
Db 5089 MSSCRSRAHASSAGNAGPKWMPSETKLSSAPRSVRGFPFSRFTGRSSRSALIPKR 5148
QY 162 ----- 161
Db 5149 MLHSRAXRRTRDSTRSLHSEKFTITRAGCSFLSAMRRDVKPELSVGSMSKKVLVIG 5208
QY 162 ----- 161
Db 5209 GGPAGSTASLLSKAGMSVKLLERETFPYRHIGESIASSCRTIIDLVGALEEVDKGYTV 5268
QY 162 ----- 161

Db 5269 KNGVLLRMGKEDWAIDWPXIFGPDVRSWOVDDPDHVLKNAVKQADVTGVTVKRVL 5328
QY 162 -----DTGE----- 165
|:|:|
Db 5329 FGDRAAGAELWLPDPTGELVSEFDYVIDASGRAGVISKHLKNRPHRPFIRNVAIWGYWQ 5388
QY 166 ----- 165
Db 5389 GGSLLPSSPTGGINVISAPDGHYWIPLRGDRYSVGVFVCHQDRFLERREHADLEAMLAS 5448
QY 166 ----- 165
Db 5449 LVEESPTVRDLMANGVYQPGVRVEQDFSYVADSFHGGPGYLAGDAACFLDPLLLSTGVHLA 5508
QY 166 ----- 165
Db 5509 LYSGLMAATSVLATVNKDVTEBEAGAFYESLYRNAYQRLFTLVSGVYQQOAGKAAYFELA 5568
QY 166 ----- 165
Db 5569 DAMTPERATAEYEQVDGAVAPAEVLVAGLADIHDAVTGTHEDRAQQAHAQADALPEDNSVR 5628
QY 166 ----- 165
Db 5629 QLLAAAEANARLMAAAGMPAAPVSEAPGKMDAHDLYDPATGLYLVTSPTLIGIGRSRXPQR 5688
QY 166 -----OGC----- 168
|:|
Db 5689 XCRASPLGGEYGARHRSVRGCXVRYXPSNASDRLARRGGPGAGVVEHTLPALPTLQ 5748
QY 169 ----- 168
Db 5749 GDGPPVEFGPCDIQPVALPVEGSPPLPAHRDGAAGHTGHDLOQAORSGHARADVHH 5808
QY 169 ----- 168
Db 5809 CGQPAADPAAQREQEPGDVDVDIVPDHMRIAEHPHGLAPGGRQVVRQHTPEAQALAS 5868
QY 169 ----- 168
Db 5869 GTVDIAQPRDHTDGIRVPAQLAFABELGVVVDGVGPGRAGLPVRRFDLGRVDRPARAQ 5928
QY 169 ----- 168
Db 5929 YEHRVSRSGORLQHLDOQSAFGLDQVGTPTFGEDRTEAGGQEVDDVGLRQSRGPAADHR 5988
QY 169 ----- 168
Db 5989 RERPVEGGRRARAVSGBEPGRQPGSDETRRPDHRRTPVGOQVRRGHSAPRRARGDAARSAR 6048
QY 169 ----- 168
Db 6049 EATRGTSPPGGTVTARRARXARVRSGWTPGGLRXPSSGSTNTAGQPRERAGATSWSTEL 6108
QY 169 ----- 168
Db 6109 PTTSSSPGGRPSRARMPSSRSGSLATPSKQSRITTVSRRTAVSPHSASTSAHSGXPATTV 6168
QY 169 ----- 168
Db 6169 VGISRARAVTAIEPGISWASSRQARRTRTQLSSAAABERMPPGRSAGVPPSRXKRSVSSM 6228
QY 169 ----- 168
Db 6229 KSRRVTILGSAVTARSKVPXASQASSKSSXXXWTPPTPSNAIALCPVIATRSRILHTSRREY 6288
QY 169 ----- 168
Db 6289 TRINGSAARRPWTHRYROESSQFTILRLKPMPSGEEIRVHLKCLTKGAAVWGPWKXST 6348
QY 169 -----GNFTT----- 173
|:|:|

Db 6349 RSTAQSPRTDHRCRGSWTTRSWGCAVSTIACSPSTPTSLRATTSTSPAARIPTGRPG 6408
QY 174 -----FNMFFNY----- 181
Db 6409 YRSCARRNSSPSWSRVRISSGAGRSPRSCWTSRRTSTSMGTTPXSRRTTARQRPQTRT 6468
QY 182 -----PRWN----- 185
Db 6469 RATWTRAGAGAGAYGPRWTRPWPWTAACMCPARTSVPCCTGTSTTTTIVSGWXPTRW 6528
QY 186 -----TKLYVGPTK----- 194
Db 6529 TRPPGCPSPRVPWCTPCALRITRAPISPTRAGPMCWSSWGRPKRSPTPSRAFGWTR 6588
QY 195 ----- 194
Db 6589 TDDVPRGRSTRSGPRELTSRLPAACGTAPHTVRRSSHGPTTGPAGHPRRMPAALPG 6648
QY 195 ----- 194
Db 6649 SGVLRARSRCRIPRLRRPARSRPGSRCPCPRECRSRPRPPSHPARTRCPTARXP 6708
QY 195 ----- 194
Db 6709 PGTPAPVRVRRAAGRAAARPPHRSARRSRTGEXSRSTPAPRRAPAPGACRRRARTR 6768
QY 195 ----- 194
Db 6769 PPAPPCAAPAPYXAGSPWPQPRRTGAPGRPGSGSRARRRRPRSGRAAPGNCRYSP 6828
QY 195 ----- 194
Db 6829 RGXPAGWXSAGXXRCPAARPGEXHRCGRGPRRRPSAATARSPPAPRRSAPGV 6888
QY 195 ----- 194
Db 6889 RPPCCHPAGRPRGRVPAARPSGWRPRTGCRRTIPVGVRSSRHRSPGACXPAAPSTAS 6948
QY 195 ----- 194
Db 6949 SGRPLRAAPAAQAQAVVPAAGPHGRHPVGEPAQRLDGLPRSGRGRDAADDAAEVHRH 7008
QY 195 ----- 194
Db 7009 EPAPGPGGLRRVPVDVGEVVAVAPHVDPAVRAGHRPAAAPVLDQAQVOVQCHPPARP 7068
QY 195 ----- 194
Db 7069 AGPAAQVGLLVGVQHLVEQDLAQHRGPGEEERTLHIGVLQRPWGPPLRQPRRGKMDP 7128
QY 195 ----- 194
Db 7129 VAGEGVLLAQPGQVRVRLVELVERGVGVVAVLQRAADAVLRFPLQPGESAHQVGRPG 7188
QY 195 VNVDSQTIYPLGL----- 207
Db 7189 VGVDEEHWVRAGLQGAADADVVRGEGIAVAVADRHRGPGRAHMLDTAVRRVVVDVADL 7248
QY 208 ----- 207
Db 7249 QVEAGVAVOQGAETPVQPAQLVGPDPGRDHPRTLGIHLLSRSSOGLVVRHFRFTARIA 7308
QY 208 ----- 207
Db 7309 SVSWAAGSNWSSSTAXGMTAPSGRVTGRRAVNRCSNAVASTSPNPHIRLASSMTT 7368
QY 208 -----TALLRYAQNCNTH----- 221
Db 7369 SRPVFSADARIVSQSTGKIVRGSTTSIDASSAMASATARATCTHPVATTVTSSPVWRIR 7428
QY 222 ----- 221
Db 7429 ADPSGTGPSPETSTVPIALGAKTATGSGSRSTAESSRXPASEAVAGRVTLMPTCANRL 7488

QY 222 -----SFY----- 224
Db 7489 XGWECPAMPEREPEPGTRXVTGTSTCPVISENLCAWLAI CSNTRXSSEGIWYSTGR 7548
QY 225 -----LYNAMS----- 230
Db 7549 CPAIAAPVARLVKACSDRGVSTTRPSPKRSCRPLATWVRPIRTSPKTOTRSSRAISSCS 7608
QY 231 ----- 230
Db 7609 AMFSAVMYASAAIRSSLRAVSGCRRHKGAAPRGGRPGRGTPRPRRSACHSRRRWR 7668
QY 231 ----- 230
Db 7669 PSHRPRGSCGAAGAHATERPRRRPGSSAPPAPRRSGGRXTGRCGSPGRWDRRPGSVR 7728
QY 231 -----RNLFRVPKYING----- 242
Db 7729 RRRPRRRRRRPPFRPPRAPSGRTRPPSPGRRPAPGPAAGWSRSCSRRTTAPAAATGP 7788
QY 243 -----TKLNTMKLRKQ----- 256
Db 7789 PGXTPAPRPGXAHRRRRPPPPGRHRAAGTTTPHRRRWGRPPPPRCWRHRRRGXRRRA 7848
QY 257 ----- 256
Db 7849 SSRPGSGRRSCRTARPASASGLPLWRGHRGDCWXRRRRXAPRRPRPPPGRCR 7908
QY 257 ----- 256
Db 7909 GGCRRSPRTATSPGSLVLRIPXSGAWSGRAGRAVERSSPVSLLPWSSAERPVAHRQRH 7968
QY 257 ----- 256
Db 7969 EHQOQYQPGDQLDRVLGDPGAARDARIAEQLPAGHRGGQGVPLDGAQPAHLLRRD 8028
QY 257 -----APVKEQFEKK----- 266
Db 8029 EGVQRHQOQORHRRPGGLRGQOQOQVDPDPDERVAPQOQOQERRQRLAGALVPAPADQ 8088
QY 267 ----- 266
Db 8089 QSAAEEDRELDQAEHVGEAAAQORDARLRHRAEPVGHAPGDRAGDQGHAGDTEHRL 8148
QY 267 ----- 266
Db 8149 SEQPRHQELAVTAARHGHAHEVPEDRHREHHAIVEDQPGIAPPVHVGAADHPAVG 8208
QY 267 -----AKKTQSTTTP----- 276
Db 8209 DGPAGPSQHGHTGSLRPARLVACRNTSSSVGRSDTSTTTPWPCPRSSCTISTRSPTR 8268
QY 277 ----- 276
Db 8269 SRTGASIRRASASTSGSVPVKBRATXAAADRSSVAARVTTIRSPTRDFRVSGVPSTTPR 8328
QY 277 ----- 276
Db 8329 PASITEINSAKSASRYCVSVSTVPCSTIERTIPTSLRLGRSPVVGSSRKSLGTIV 8388
QY 277 ----- 276
Db 8389 KNPAARSSLRRIPPEXCLSLSAASVRSNCSSSSVRRCASLRPTWKRPHISRPRSPVR 8448
QY 277 -----YFSYT-----TSAALNVT----- 290
Db 8449 ISSTAGYWPVTWFRRTSAASRVTSRPTARAPPGLSLRMLRMRTAVLPAPLGSGRAHT 8508
QY 291 ----- 290
Db 8509 VPSGTTNRSPNSAVSDPXVLRSPSATRAAFADPLILLRRSVRTGQRRSXTRSAPADGRR 8568

QY 291 ----- 290
Db 8569 ACSSPGRTRTRVRPARRSPVASPRPEDWRPTGRRATARHWFGRGPPRGRAAGPAVP 8628
QY 291 ----- 290
Db 8629 TGSAAAGRGATGSPRSWHGVRSRRRPTSRAGGCARRPRRRRTTRRRSRXSAGSGRR 8688
QY 291 ----- 290
Db 8689 PSPAAPXGMSCCRFRPRCRSRAGPASRSGRAAXRARRPRSTAAGXGGRRAASPTAP 8748
QY 291 ----- 290
Db 8749 RPPGTGRSPRTTISGASWPSRSXPGORTSRTAGCCRTGCRGRGTPTSRGRGYVYP 8808
QY 291 ----- 290
Db 8809 DRESPPCRRTRARXCSAGRTSRPGPRGSSSTRTRARRPASPAGSAAACWGCXTPRSALXP 8868
QY 291 ----- 290
Db 8869 HRDCPVRSWARSRQRREXAXRPAPPRIARTARCTGTSRPHSGRRHRRWRRTGSGTAAH 8928
QY 291 ----- 290
Db 8929 TWWDWAGASYRVPRLRLRLARLAGQODVVERLGRALDALAGEDPVEVVVDEARAQP 8988
QY 291 ----- 290
Db 8989 VGQGGPQQPDRTARRPAVLPAVVAQEQOTARSAGELDRSVHGEFPLGLGRESVGGP 9048
QY 291 ----- 290
Db 9049 APQVLPVLLPDMRGQHOPARPQQFERLAQMQRVHRVFDLGGEHTEVERAVROAAQP 9108
QY 291 ----- 290
Db 9109 GLVQLVIRROVLAALHLRLQPGAGEQVAVRGARSEVEPAPRRPGELLQORVAEVEQMG 9168
QY 291 ----- 290
Db 9169 VGPVAGARIGLPEALGDDLLGGGLAARARRRRGERAQQAQVHHROYSSCGNRPE 9228
QY 291 ----- 290
Db 9229 IAGDSQSRYCSTGRYSKSSSRCSPPAARCSRHARSTCCLSTYSYSGTCGSPXSTSPGR 9288
QY 291 ----- 290
Db 9289 YSXXYTLRLCLPTTAGAFCSVRMLCSRASPGRITAGTAPVASHVALSRSAVSVTCSSQX 9348
QY 291 ----- 290
Db 9349 LDCGMAQTLSSGAWSEXMTSMTRPPVTATASEYIWPSCQASGAPAVLNTMLSQV 9408
QY 291 ----- 300
Db 9409 GMKEGPTRVSMARSSSGWSPASRAITGDLSTMYSMTGASGWSGVQVMSCHMDRVQAS 9468
QY 301 ----- 300
Db 9469 SIRPAAASCTDSSAGLSRPRSTAPTCRRKAVTCGSRPVSRVSIWHSGTAREAGAVSV 9528
QY 301 ----- 309
Db 9529 MSNALLGGNGRTGARDRRRSTTACAARRRRPGRGPAPPPASPAGCTACSGRRPPAAPT 9588
QY 310 ----- YRPD 313
Db 9589 GRRWCKRSRAGRRRRVRRPRGTAGAAAAPPRAGRPPRRPPRPXPGPRGRGRRPD 9648
QY 314 ----- 313

Db 9649 VRCPPRSLPGSRRRGRTDGAGRCRGRWPRGRSPCQGWRRPRRPAIPXPRGSRGAAVRR 9708
QY 314 ----- 313
Db 9709 TAAARPRSAGPAGRGACAXGSRRRRHRCGSRGRRRPVXSRSTGTRTSRAGGRAS 9768
QY 314 ----- 313
Db 9769 PWPQTDLSRAGGWSPLPSSSGASASSSSRRASSSRSTARTLLSSRCSGSWAPMTG 9828
QY 314 ----- 325
Db 9829 MTEGWASTQASAWATLCPRSAANSFSRSTAASSLAKRSPWKVGEKAROSSGGNTWPR 9888
QY 326 ----- 325
Db 9889 SAPASTPCASADHPMTPIPAAAQCGSTCFSPRSMMLNGTWTATCGVTTEENSAISSTVKL 9948
QY 326 ----- 325
Db 9949 ETPTYRASPSSTNPASAPAVSAGGTGSGQWTRNTSTXSPERRLSVRLARRKPASLLR 10008
QY 326 ----- 325
Db 10009 TTPSPRSSRPPPEVENTMSERRVRSRPRPASAPASAPXACAVSKKLTFRSNARRTAS 10068
QY 326 ----- 325
Db 10069 TAGSVRLPONFLSAQVPROTBETSGPCRPVSRYLTQPPVPCAGXGSRRRPGRRATRVFS 10128
QY 326 ----- 331
Db 10129 ARPPPRACGPRRCRARTXIPXRRPGAPRGSSXRTGTAAWXPRRSPGSGXRGTRRAAAPA 10188
QY 332 ----- 331
Db 10189 GSRAGRPRRRSGKRXARIRRRRRRPEVRPGSRGCGGRXWAAPIHRWGSFGTGTG 10248
QY 332 ----- 331
Db 10249 XRPRTAGCPGAPARGPXPXRCRSTDRCPGSGTAPRPTGSRARHGRADTRHGADSPRA 10308
QY 332 ----- 331
Db 10309 APACPRGRPTPRPXVPRRSGASAPPCRPVSGLSCSRSTPDVARTVHAVVGHVVRVGALEPP 10368
QY 332 ----- 331
Db 10369 EQGALLEPGERHLGLPAQLRPGGLDVGAAPAGLVLRRAELQPQLHSRECGDLRLDLDG 10428
QY 332 ----- 331
Db 10429 HLTGRADVDRAREGRVHQHQSPHRVLDVAORPGLGMAAVDGERQSGERLVDGRHHPV 10488
QY 332 ----- 335
Db 10489 VAAHPRSVGEQAGNAGVRAQARVGGQORLGEFPGLVVHRTRADRVQVAPVALGLRVLG 10548
QY 336 ----- 340
Db 10549 GIPVDLUGGDHEAGAVRHQRQVRLGPGRADPQRLQORQIVQGGRRRQVDDVVDRLP 10608
QY 341 ----- 340
Db 10609 DGERLGHVGPQOREPLLVGVERVEVRDPAGGVVHAGHROTPLDEPPAQVGADEAGTGGDD 10668
QY 341 ----- 340
Db 10669 GPLSLLOPCHALLRASVALSTTSATSFADSPSPATRRVSSPATAAGEGARSTPSPSFA 10728
QY 341 ----- 340

Db	10729	ARRXWSNSTRSASSTFSAVASRDRVRVAPAGSIRRPQISVSXLRARDSCGIPATMPF	10788
Qy	341	-----	340
Db	10789	MXQLALPWCTSRQYGRMSSGQLRSTRTPAGSAPSSSTCSGAVSTTSAPCHTAASMAC	10848
Qy	341	-----	340
Db	10849	RSAGTSAPKMEAEALPQTMHLRPLRIGPSCHGSAPWAPLXGTXRMGRSASPSGGTATSG	10908
Qy	341	-----	340
Db	10909	EGSTAYIRSRPHSTPYFSKHDGSAETRSSPGSVSTVPMWPGDTYTLGTTPRSAAATSAPS	10968
Qy	341	-----	340
Db	10969	TAMMCTTRSGROWSPKLTALSXLCTASIDFPQKSRSPASKLSLSDCIGCPWNGCSGS	11028
Qy	341	-----	340
Db	11029	RGPDCGVVMAC TXNASRALSIMGASSTGTGIRPVAATPRSCDGHATRISCPAARSQA	11088
Qy	341	-----	340
Db	11089	ARGTMMXCPAOEAVNIILIRKTPXAVRCLPGARRRSAQAGAGAGAYSMTWVRQL	11148
Qy	341	-----NPFCEP--	346
Db	11149	PLAFGEPMWTSCTTTSPLWLTGTTGTPVARTPATSSASGRCDPGTASRHLPGPAS	11208
Qy	347	-----	346
Db	11209	SROMLQTVRGPTLRSGHGTSAARYGLRSTAXAVGTGTPSTNAASCASSRCSRS	11268
Qy	347	-----	346
Db	11269	RSQAQCFKLCRTRXRTGRSSPSTGKFXSSERRPSSVKRMSRSAPALRAIASGARTG	11328
Qy	347	-----	346
Db	11329	PAFAKPSRRNRSTAAARSSVAVVMDASSFLVRAASRAGHAGLACRDDVADLCRGPLVQ	11388
Qy	347	-----	346
Db	11389	HLDIGEARLLQGAHQRRPVQAHVDAEAAAGLGLGVFADGVDRVPSRSVIPSELPOXWVE	11448
Qy	347	-----	346
Db	11449	RXRQSETEPMSASSQRCCTWXRSPAQSVASKWPSPANISDTSASRSSTLVTPCTSMTN	11508
Qy	347	-----SRNRTAVERSEFMKN--	359
Db	11509	SARRPSSRPSRARISTPGCPDIARVAPSSAPSSRTPLAPGNAKXASARSGWGLSRX	11568
Qy	360	-----	359
Db	11569	SRVEAVARTPLYTKSFTHGSSSTRRRISRSFTVALPPLSGALPSETAWAGDHRSAPLS	11628
Qy	360	-----TH--	361
Db	11629	RSATHCAHSGTRSPPIISKEERSAGTEAAHAAVRAASPSAGVSAQSPKTSRASSVAV	11688
Qy	362	-----VLIRNE--	367
Db	11689	XATITSRAGRCAMGLWGTFAAARTAWASSSTVLVRDEPTLKARPOAFSVSGARTWTSTM	11748
Qy	368	-----TPYTI--	372
Db	11749	SLTXLNARTXPPSPXTEIGSPRIWLKWMATALXGSRMFCHSPYTLICIRNAVKGPRWVR	11808
Qy	373	-----	372
Db	11809	IVSSSRSTVSFALPXLISAGTGIDDSRIGVSPCYTATEDAKQKNETPVRSAAALMFTLP	11868
Qy	373	-----	372
Db	11869	ITLVPXLRLFTEKLIASAANAANKWTWNRCSAKSLSTSVTDVPNRCAPSGTRSAPV	11928
Qy	373	-----	372
Db	11929	PLRSSRTDTRCRLSRASATCEPMKPAPVPTVQCAIKSPGSKGKQRETKNXSVSEGEARGR	12048
Qy	373	-----	372
Db	11989	ALPXSUWLSARPEPCERTABARSTAPSPGSRRTSARCRGLARTTHGVSRARRGRRRT	12048
Qy	373	-----	372
Db	12049	ASSAGRVPSGRVHARPGSRRRPGRPVAVASVPDWSARSPRCGRPAPAAARVRRPAGCPAR	12108
Qy	373	-----	372
Db	12109	GRTRAPSGRCRHRPRTAGCCGRPAAGWXTAERPSPGTDRARSCCRTASRSPAAGCGRRR	12168
Qy	373	-----	372
Db	12169	GRPCRRPWTTPRASWGRSARPRPSRSCCCTPRRWRPARSWPCRPCTPAXVPCRPAX	12228
Qy	373	-----	372
Db	12229	WYVXTAPRRRSTGRRAIRPRSXPPDRGGRPPRRAAWHRRRCRRARCTGRRPAPGSPGSR	12288
Qy	373	-----	372
Db	12289	RSACPARSPAPDRRFPVRAARTGSRXSPLHXPGRVAVPHPRCLACSRKPPSPSRPPP	12348
Qy	373	-----	372
Db	12349	POACSTADLDPAVVSQHQPVRGRRPGGVHLGLPADHRVLDVTGDSLDRSATQHDRVLDLA	12408
Qy	373	-----	372
Db	12409	VRDPAVALDGGEGTEVGLHHRVRADGGTDAGAGDLRPLRHDSAHQFAGLVHLAVDA	12468
Qy	373	-----	372
Db	12469	GLQGLQQVPHLQVGDARDVLVPVDHHRGADLVVVGQPLQGVGDQLVLPVPPRDPAGHG	12528
Qy	373	-----	372
Db	12529	LVHGRAEGVHADDRQVAPCVLGLLHHRGHPAVLVKFGHAEAPARVGDPGEHDLRGGLVAAB	12588
Qy	373	-----YGTLL	376
Db	12589	VRHEVDDAVHDVVVAQVHDEVVVAEBVTGHEHGVGPARGVLLDVGDVQAEALRAVSHGSL	12648
Qy	377	DMS--	379
Db	12649	DLRRGRPDNDPLRYSGLLDGIQYIGENSSIGHRIDQLFGRGMNRPPQPRANPAHQYERLH	12708
Qy	380	-----	379
Db	12709	IPHLAVASVSRGPKGTPDGIETLPTTIVLACPSGFPVXSGPAGAAALSPPRPLLTXYVTE	12768
Qy	380	-----	379
Db	12769	SVRSGMAVRRVRFREAFQVDDVADGGQLPEDVGRHVPAEALLQPAADHRRHHRVEPL	12828
Qy	380	-----	379
Db	12829	LHORAANVQGRGESHLLDLHLVHEHPGTQLRRQTRERHHVRVVRVAARLVGVRGRGL	12888
Qy	380	-----	379
Db	12889	LDHRVAQAAQARHRASLPVRDQGVHGVGDAGIERDLDRRRLRWRAVDGDLVDVHDHAEUR	12948

QY 380 -----SLXYNET----- 386
Db 12949 EAHEGRRDRHHGVQGRGVRGRLHSVQYTTETGGQSLDAPETGRGYAHLGQPPGQ 13008
QY 387 ----- 386
Db 13009 AGVGRPEPVLARVQQRPRQTGEVRRGRPGLGMDPEPGVGVGVDRPGGQPHILALADD 13068
QY 387 ----- 386
Db 13069 L0LPLTRQHRVHRHVEGAALFALGAGROERDRQHGGDDLHAVEGVAARVAARVAVVQ 13128
QY 387 ----- 386
Db 13129 RAVPQPOVAARDAAGRAQQRVGVDVVAALGRSRPGDTPQRAVLPGVGRQDFEVPVRLER 13188
QY 387 ----- 386
Db 13189 RPVHAHAVAGGAHGAEGPGLGLVAPXRSDHVLHAAFQGRLVDRVRDHGVRGOLDER 13248
QY 387 ----- 386
Db 13249 PVAVLGRRGRGCEPHRVAQVRPRVAVGVGHRGPRVRRERRVRCAGTQLVGGRAQP 13308
QY 387 ----- 386
Db 13309 VEQGVHRAVEGHVGGDPYDQPLVGPAADLLDLAGVTGDHGRRRGRGRDHRVIARE 13368
QY 387 ----- 386
Db 13369 PLGGLVEGQDHRHRATPRGLAQQRPTADQPGVGHGQRPGGDGGDLTDRVADDDGGL 13428
QY 387 -----MFVEN----- 391
Db 13429 DAPXPPQGEPELHREDHGLGAVDLGVVALQQHLPGREPDLLLEHRLQFVDHILGEDRLGF 13488
QY 392 ----- 391
Db 13489 QQFAAHPGLPGVTREHPDGTAGIRQGPQDEGTGRFLSQRAQPGGQFVPAGRHDGRPV 13548
QY 392 ----- 391
Db 13549 DQRLGPMQGVQPDQGVQVALLDPLGERGRAVTERLRGGRHGRGSGRLRRGAGRGRL 13608
QY 392 ----- 391
Db 13609 FEDHMGHGAVAETGHSRPAQVETLGRQPGRQPDVEAGPVDLGVVRVAVQQAGRDPVLV 13668
QY 392 ----- 391
Db 13669 EGEDRLDDPGDPCGPPLQMPDVGLHRPDRARLLTGAVTSVDGGDRADLHGVAEAGTGAVGL 13728
QY 392 ----- 391
Db 13729 DEVNGVSDARGQAGPHDVLGLPAGGHDVAGPVLVERAAAHQRHPVTVAYGRVQL 13788
QY 392 ----- 391
Db 13789 DDQHAALAAAVAVERRNGVAGSRGRGERLAPAVGRQAGRAGOGDRLARXHQQVDPAGQR 13848
QY 392 ----- 391
Db 13849 HGALAEPOALRGVQRHOGGRARGVDEAGPAQVERVRDVGHAHRADPRPCVDVRQV 13908
QY 392 ----- 391
Db 13909 FQPSGAIWVADPGEHTGVPCQGVRRDARVLQCLPGHLOHALLRVHQLGLARRDAEEL 13968
QY 392 ----- 391
Db 13969 GVTRYVVVQETAGANRRPORRIEPRTEVPVLPFVAVGGHLADGVASGRQVQVPEFAGRVD 14028
QY 392 ----- 391

Db 14029 TGEPASDPDHGHRREGVRAIKGYGHIFPVVAGIHSSQSGDSRTKRTALDTPTRGAVA 14088
QY 392 ----- 391
Db 14089 SCRTWSPGASSFVRTMCSMDSRWSLAPQADLPAEFPPGQGRHPPGRHSALLDHFAXXF 14148
QY 392 ----- 391
Db 14149 MKETRATSAIPMPAPARARSATSIWSXISAIATKAGLRVANCWPIRALKSVARPRVMLG 14208
QY 392 ----- 391
Db 14209 GNNGSSPRCSDISGSPLEATEQSRAPKRLQHOGGSASPPMKPGPIRHSISGVPLXSA 14268
QY 392 -----KTASDSNKTP----- 402
Db 14269 ATISALVRFSGACRXPPTASSSALNTPLOCLISRAMSPPTRCRCRYAWGGQSPFWCQV 14328
QY 403 -----TSPSM----- 407
Db 14329 LAYSSRSNSVGDPPHLFTACCAARKAISGGGTSPAMATPPSSCASASAGXKRPRSTAHT 14388
QY 408 ----- 407
Db 14389 LTHGGTKTPRPMYRSAICTNPPPCFQAYWTRRARICILLTIIVSLVSLFRESRPAPA 14448
QY 408 ----- 407
Db 14449 FPGQNSPHRRORTTELAFRAGAVRASARSSSGAVMTVAHRVTMTSMVNNAGVRTPS 14508
QY 408 ----- 407
Db 14509 SSPWRIITSSVSPQFMSTPMERRASTRSTDISAAAAALPASLPTTATSRATPAHTRAPP 14568
QY 408 ----- 407
Db 14569 PRPSCRSVRPTEAKNGNSTAVTSSCSLPRASARAGAVFGOMAPNRKAPORWMPSS 14628
QY 408 ----- 407
Db 14629 VATADSRTPGMISVSTPPSGRRSRSTRRKAGTSTMTPKKTAVSARVSTPRGPP 14688
QY 408 ----- 407
Db 14689 RNRATRPASSQQARSAVAAPTMAXVPARDWSIPRSISSRASTGKAVTLMAMPKRKVRNGR 14748
QY 408 ----- 407
Db 14749 KPTSGAXFPVYSAYPRAQPRANDTQPTIPTAAPVRPAPSMWVSPNSMPIRMNNSASPIW 14808
QY 408 ----- 407
Db 14809 ATRSRXEWRAKRASTWTPGATRPSTEGPRTIPEIISPTTRGWPKCRPTSAPQARATAT 14868
QY 408 ----- 407
Db 14869 TARASSNRHASTCGASPGVALPEPGRDASSLPVWPCHSSRAAASGSDSRHTASTRVVP 14928
QY 408 -----GFQRTFIDP----- 416
Db 14929 RDVTSATRSHNFQGRQKQAAASGGQEVAFPGDHRVPVDPREDQELVGLKRPGLLIGDDR 14988
QY 417 ----- 416
Db 14989 DVRAGRVAELVLVHLGDAGQERLVQAAVLQDDVALRGAQVAGVQPLVPRGGBEFQESV 15048
QY 417 ----- 416
Db 15049 TGLLGPGBELRVQVRLSGADQORRLVDLARGAAAGQHEREASAVDGGQLGHHHLQAGA 15108
QY 417 ----- 416

Db	15109	LGHPDLAQPVVVQVLVADHVVGAGVEHRCQIALLLGHPDAVVRQQAADLPHELHGTRQVV	15168
Qy	417	-----	416
Db	15169	EHRDRGDDLGAAPAEQVVCBEVGDQLDARRVGRGEDGGGVHSDQAGPVGVRQOQAVV	15228
Qy	417	-----	416
Db	15229	AADVHQVAGAEVDQAGDGDGLAAQVVRHGAVEAGPVAVALAVEAGVGAQLHOFATA	15288
Qy	417	-----	416
Db	15289	GRAAHQLQRNVHVLAAQPREHAGEGLRSEVEHQQLGVVADPATVDDGRVTGRAHDSPV	15348
Qy	417	-----	416
Db	15349	SARLRLSQPMTSAASQSGSLTRGNWSSLRAPAKTEREPTASLISRLLTVVRAPTSHSP	15408
Qy	417	----- -LW- 	418
Db	15409	SASRWEATOPPHSTGAKLRDSPCAXTGCSRMSMTPLVLLSPYMGPRRTTRWPAQSS	15468
Qy	419	-----	418
Db	15469	SASARFPXGVSSWVRAEMVLIWTKRSHSQRRSARVLSTVSRSSSGCCRAQRRREGQA	15528
Qy	419	-----	418
Db	15529	ALTSTSGPLRRASSXLARVTGMAEPSARVAVTPNPPAPTATLTGAPACAAAARPEP	15588
Qy	419	-----	418
Db	15589	AADRIIPRAWXWRXTSGTRWGRSAXPRPASWCRPRARGXIPLCSWGRLRRPWXSARP	15648
Qy	419	-----	418
Db	15649	SPRRTAGTROSAGDNAGRPFGTGAAXYIPGWAPGGRKXFRRAPAPYPGYRSAAAH	15708
Qy	419	-----	418
Db	15709	FSCELHSPHATDSQASIRBEARVRKFLVLSAAPMIRVPSADNTAXRPASPAGQMPVTRG	15768
Qy	419	-----	418
Db	15769	FAAGTGTPOGRMPLHPGAPQLPSPRRSLRTAMIVKSGQAPGPRNPPLHLEQNPQEP	15828
Qy	419	-----	418
Db	15829	PAPERKSTTAGERVLTGKESAAGAWSRPGALDTSQHSPPREYQVTRAPTTLSSLDAKIIE	15888
Qy	419	-----	418
Db	15889	RIAGERTTICASKLYLSRQVEYHVSKMMEFRVFNRTALVAKAYALGILRSEVWPPRV	15948
Qy	419	-----	418
Db	15949	LOBRVNKATVTGSGRLRRAGQNAHVANAKVYRNGFSYIPORSVDYGMAGRPQVTR	16008
Qy	419	-----	418
Db	16009	PVSGQARRDAGPFEVHRRRLGAVAGPGPVAFLPAXLPCTHRAALSAALPTPRALLT	16068
Qy	419	----- DYLDSLL- 	425
Db	16069	TAGGTAPQPRCRDKWIKMSNVQDPFVAVVWVSPSRHSSVAPFGALILDASDDLGHHA	16128
Qy	426	----- -FLDEI- 	430
Db	16129	RTTYSYGGQTLPSYGCDSGSDGRFLHEDQFPWLEGLSQVHLLSDRELQVFLLSHGYS	16188
Qy	431	-----	430
Db	16189	NREIAALLGVTERTVKAHMAQILAKLGVESRLQAGLVAFAHGULPGHIGDAEGSDHPDPTS	16248
Qy	431	-----	430
Db	16249	AGSXIRSAIRGSGSQDRTRSGTSTSPRWKAAASRCPSSLKCSGSGVYRGSTYRMAVSG	16308
Qy	431	-----	430
Db	16309	PARLGPMTXWTFSSPGARSRTTTSRVCSRTSVKDRXSPSPSSGSCQEGEXMWDAMTSGVEGC	16368
Qy	431	-----	430
Db	16369	RSMIRTSPTFRGXVNVWFRAPAPXMPMTMHVIRSAGRSRVRVAKSGWRXQAPATASEV	16428
Qy	431	-----	430
Db	16429	RSPSSVATVAVQVRVGDAPLPCDSEAPXWQRGAVVWSRPTSAPAGNSTPCKTNWFRP	16488
Qy	431	-----	430
Db	16489	GHSSTRSLPPGSAPTRHVWTSRVTRGLATGLPSTRTSTCAVSGGPGISGAGRAAYR	16548
Qy	431	-----	430
Db	16549	NASPSASVRKVTPTVGIAPPTPPSGKNCRSXKGIROGAIIPSTSWRETAPRXVSTGS	16608
Qy	431	-----	430
Db	16609	FARMRRPASNIPLHSSRSGRPTLRDXQRCGRSRRFSTFLRKQFVRXMTTDTRECPRRR	16668
Qy	431	-----	430
Db	16669	NAAILEXLMTRKRLRTGTPPGVQWVWWSVPSVWRTSAGICGTSTSTXTVNTSCCWGTR	16728
Qy	431	-----	430
Db	16729	XHLRPRRSPRTRPGRCGSGWPGREGRRPVTRCCNGPTGCCSWTTTCGRWPPPTP	16788
Qy	431	----- -RNFSLR- 	436
Db	16789	SAPRSAPGRTWSVRGCDRESSAGCPGSCPGSSTAGTATAGTASRGAPAWASTPSR	16848
Qy	437	-----	436
Db	16849	AARASPTWTXAAPASTCSPVTRTPSSRXRRRVRASCILLPEHAVHVDVDPGLRLRYLL	16908
Qy	437	-----	436
Db	16909	RRAYWGAFFGATAPVGGGTPOGADPAPHRAGVALHTTAVLRVGRHRRRRRPTPAAPA	16968
Qy	437	-----	436
Db	16969	QVSHACCRTEPERTQAXVAAMPFFSGTDVQVPSAATRVTSRVLCWISPRRGAKRGGR	17028
Qy	437	----- -SPTYV- 	441
Db	17029	SRPEAAAMWSRTSVLPLPVSVDVDPDAGTPYVGNATNAATSTPTXSRVCEPSPNTVGT	17088
Qy	442	-----	441
Db	17089	PFSEASQXWAMPASPAPGLCRGPYTLPSRNTLCSVPCSLNRPMTYFSAQYLAVPXGAPG	17148
Qy	442	-----	441
Db	17149	LGSVSVFVGTGPSKXSAAPPVEANTTPTACAAASSTLIVPSTLSTRAGVATETRTSI	17208
Qy	442	-----	441
Db	17209	WAARWMTSCRADRTASSRAAAASVTSTRMTARGSRSSGSPVVRSSSTVTTSAPASRKAST	17268
Qy	442	-----	441
Db	17269	TCEPMKPAPVTSRRVPAENPGNMDMARPPSSQLRHSAAIVGRGTGKTSIWAIVAPTATA	17328

```
QY 442 ----- 441
Db 17329 RLHLGRPTAEAPVTCVSGSMELPRPGPRHYDRFAATEAVPKYWGVDXRVTATPWSG 17388
QY 442 -----NLTPE----- 447
Db 17389 STVISPAVSCSTPRRVTTPSPSAMRPRSSGSPASPATVRXRCGTRAAAVTSCXWG 17448
QY 448 ----- 447
Db 17449 CQTGARGXGPPMTWTPRWSRRRISPCSRPGXPRASNGVSRASVSPPIWRPRP 17508
QY 448 -----HRRVN----- 453
Db 17509 PACGSAXRRRAVCRAPASPTSSIRVRCPPSPGPRMWCWPTCRTCAGHALGRAGARA 17568
QY 454 ----- 453
Db 17569 AGGPGCCARSRRRCPRMXSPSPATAVRGRWLPRPWSGXRSAPARYWSGQPTCWRQAP 17628
QY 454 ----- 453
Db 17629 EQSRLNFVFSVFTSVSPVPFPTXTHSAGGYRCVLGRAWRKPGCSXPGCIFIGRDX 17688
QY 454 -----LST 456
Db 17689 SARPLPMMRRRVKRAVRSGSAARFRGWTSSRGIFRWSRTCAGEHILCSCSRPGSICRT 17748
QY 457 SNSLW----- 461
Db 17749 SRSNWPARFGTGAXXSISDGHLGAVATDSSVGTSHIENMRRLXSALSDLVLPKLTG 17808
QY 462 ----- 461
Db 17809 PLPDGAEEFFCFGMPLVHRHRELGPEDRYELOYGISNMGWKPFIEVVEAARQARLRRI 17868
QY 462 ----- 461
Db 17869 RVCGRWWDGETCFGFEITSAGAGTAEASRSRRCRSDMSSPRDGSPSSRRCSARSWP 17928
QY 462 ----- 461
Db 17929 TRGCSRPGCSKRRRAPCCRCPRKNSCPPCTRVTSLRXWARTPLPRWAVMOATSSGTR 17988
QY 462 ----- 461
Db 17989 RSWGSGRTGRGPGTTTPGSGSCGSSADPLPYRFVEDRGRGVRRVQRARGAGHRDADV 18048
QY 462 ----- 461
Db 18049 VAQRTPGVAEPGCLADQQQGLGEVEVEHIAVALVRLTDQDERRAGGPAGLHPVHHLAR 18108
QY 462 ----- 461
Db 18109 XRHPDGGQGEQAGARPDRFRVVDVGVGPRHDQFVRAQIRRANDAADVPTGRPVEQVA 18168
QY 462 ----- 461
Db 18169 QEVLPDRDPVEVVGHLHDGDDLGRGLVLLPOLHQVGRHRDLLRPHRTKLLPRPLRQLP 18228
QY 462 ----- 461
Db 18229 LSVVEERAYGPAVLHRRERADSHHQELTGARALVAVEOCLPILLETSVAXGDPSPRRRP 18288
QY 462 ----- 461
Db 18289 LSTXSCHSKBPIFRSMHLONDCAPRLRLEAGQAAVTSHGINAPCAQHGRYRVRPPEDEPL 18348
QY 462 -----WWL 464
Db 18349 DPAWRPVDLTRAPFEDRAEDRAPWPDPRFALCWNL 18384
```

RESULT 13

ABP76679

ID ABP76679 standard; protein; 19938 AA.

XX AC ABP76679;

XX DT 26-FEB-2003 (first entry)

XX DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 3.

KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.

OS Streptomyces viridochromogenes.

XX WO200268436-A1.

XX PD 06-SEP-2002.

XX PF 24-AUG-2001; 2001WO-EP009815.

XX PR 25-FEB-2001; 2001DE-01009166.

XX PA (COMB-) COMBINATURE BIOPHARM AG.

XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;

XX DR WPI: 2003-018650/01.

XX DR N-PSDB; AB237515.

XX PT New avilamycin derivatives, useful for treatment of infections, and
XX PT nucleic acid encoding avilamycin synthesis enzymes.

XX PS Example 1; Page 68-301; 319pp; German.

XX CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (AB237515-
CC AB237516)

XX SQ Sequence 19938 AA;

Query Match 21.8%; Score 534; DB 6; Length 19938;

Best Local Similarity 1.2%; Pred.No. 0.013;

Matches 231; Conservative 105; Mismatches 126; Indels 18744; Gaps 79;

QY 1 MGRK-----EMVVRDV----- 11

Db 615 LGNKLANIRESDAICQVIRAFKDNVHVHVGKSPKDDIETINTELILADLTQTEKVLP 674

QY 12 -----PK-----MFVLISISFLL 24

Db 675 RLQKDSRIKKLAPKPKVAVEAKEILRGDTLFAHGIVGQTERNELLDHLLTTKPFY 734

QY 25 V-----SPINCKV----- 32

Db 735 VFNVDELDIDEFKNEQALVAPAEAFINAKLEADLAELDDDEALELLOSVGQDEPGL 794

QY 33 ----- 32

Db 795 ATLARVGFTNLGLQVLTAGPKESRAWTIKKGATAPEAAAGVIHTDPQGFKAIEVISFTD 854

QY 33 ----- 32

Db 855 LVETGSVAEARAKGARMGKDYVMQDGDVBEFRFNVAQFHDAAFKRAQKGSTPPGG 914

QY 33 ----- 32

Db 915 GPLLFPVLGWCWESXSSISLCPAGGQEWYVIRYHLREGVESCRXTASEARLSPRXSRR 974

QY	33	-----	32	QY	62	-----	61
Db	975	STTITRSPSRSPATPYSSRPRIMORCAPARTCCALRRTPTDCSRTRTPTLPTSMCRSG	1034	Db	2055	SKPGQVSPSHHRPQTRIRRSRACLAASSTTMNGFHRPQLLPMYCSXSRSGPSSRGRTS	2114
QY	33	-----	32	QY	62	-----	61
Db	1035	SXSIRIRTLVRPREACLRGSLGLXHLVLAQERPODARPHQQAHRGQARPLHGRQTR	1094	Db	2115	WGIPKQKNSAPSGKGVNLCGCTRSLSAEXSRRQFSIEWVTELSAVATPKWPFSEID	2174
QY	33	-----	32	QY	62	-----	61
Db	1095	AEVPLAGGVAAADRRRTTPRVPGYQGDPRCPVPLLTAAQVVLPWCDGAGMPSNPMF	1154	Db	2175	YQAPVPGKRAQGDLLVRQILPGLHEHKMCSPAHVLRHKMPDOLLVHPRKRAADPDLT	2234
QY	33	-----	36	QY	62	-----	ILR 64
Db	1155	POVGALXLMFRFNVAQIHHVLLNAQVRGLVIGPLALSRASGLGVAOSSPADRAV	1214	Db	2235	AEFTRLHTGGNGLAESRLPMKIPGYEHPGRHSALPRTHLPPLALXVXGKTGTTLVK	2294
QY	37	-----	36	QY	65	QLETTISTKYNSK-----	84
Db	1215	GPCPERQRPVGRDHLAVRTGEDLGVGVAEAGQGEAVGAENAGELAGRGQVADAFERG	1274	Db	2295	TLEKTLGTFKFSRDRCSGACLOHVGCPCDOYGRAGADLOPLOGLYGSHRDLTAAVGDGHRM	2354
QY	37	-----	36	QY	85	-----	84
Db	1275	ASAYADQPGGGEAXAPLVAEEGLPLGMDRGCVVDEVTEAFVGPPGPDQDATAFOVANA	1334	Db	2355	RGORRRRAQOGPPPAARAPALPSACAPVRQVQHDIGRPGEDGGORTRIEDVRGADGA	2414
QY	37	-----	36	QY	85	-----	84
Db	1335	RAVRVLARHVALDHVAALLGLVLRHRPPLRVOIGTVDPGGDEDAADGGFGGQGP	1394	Db	2415	RORTALRRQALPPAGGLRGLQIGLAEYLLTTPPELARGQPRREOGEILLRELQRG	2474
QY	37	-----	36	QY	85	-----	84
Db	1395	GVCLPDQVFPGRAPVFPVOAGAGDAVRAGVHPPPEGVVEDHLGVVGVVDDLGEVGGRR	1454	Db	2475	VHIGGDHLPORAPVQOPQHRQEVTAARVPORHRTVAGEAGEGPLEDLGRQADGGVVT	2534
QY	37	-----	36	QY	85	-----	84
Db	1455	VRSGPRAHSITLGSASSRAIRAGQQGSGATSSTKTSWAVSRPKLPRVSPPMIVRTASRP	1514	Db	2535	RRGEHDTAGEIGTVDPPLHGVAVRTHLSTPQYFGTASVAALKRSCRCGPGRSSQLPDDT	2594
QY	37	-----	36	QY	85	-----	84
Db	1515	MSTCARAAEYASSPSSCWTRVSMGVVVRVAMTPEVGHQOAGRSRSGHARSAAAR	1574	Db	2595	QVTGASGAVGRPRXSRAGVAVTAQISEVPNVPRTIAALRCRSCBEGRAMSMPGFSAG	2654
QY	37	-----	41	QY	85	-----	84
Db	1575	SSKGARVSTGROAGSSGSSAGGRTYRRPCWAQGALIPCDVTAAWPASRRGAQSFCKC	1634	Db	2655	TRLLVTGGAGFTGSHVVDAPLEAGAEVTVLDDLTGDDPERLDPRVIRRVDTVDAALDE	2714
QY	42	-----	41	QY	85	-----	87
Db	1635	QERKIGSEFQDHSVSGRRRLGSPHATLVSSGRHCSATATSARAPVSSWXWESARSRS	1694	Db	2715	AVRSARPDVICHLLAAQIDVRVSVATPAVDARVNVVEGTINVLEAAHAHAGVAVVFASTGAL	2774
QY	42	-----	41	QY	88	-----	89
Db	1695	RWSTAGPYARSTTDSGCRSGRSGFPVRCGRSRWRPRTCXWSWGRRTPRPRSPSWR	1754	Db	2775	YGEVVPVTNEDTLPRPGAPYGTAKYCAEKYIGLFNRLHGTSHSVLRLGNVYGPSPGG	2834
QY	42	-----	45	QY	90	-----	89
Db	1755	CPPTTSGRSGRTSWAYCSTGRVGRGTSAAFPARMMPWARTGXSWRGTPPTTNRSG	1814	Db	2835	EAGVTAIYGLASEGCVTVFGDGSQTRDYVTVGVDAAFVAPYGTVPASGTSDTGKS	2894
QY	46	-----	45	QY	90	-----	89
Db	1815	RAPARCSPPGSCRRHARXWTGWRPAGPPARRSSWSVTSTATAMCSTSTSPSRPCCXS	1874	Db	2895	TVLEVLDHIAASGRDLPFRPAPRPRGEIQHSTLDVTRVAADLGWTASVPLEKIAATYA	2954
QY	46	-----	45	QY	90	-----	89
Db	1875	AMRQPGSATPGVRCATTRSASRWPAARXTRTPRPSSTKRYGSGSAEBEQQLPEQPG	1934	Db	2955	WVRSGSPVQOAXLTCAGAAAGVGRRRSRWRPRTTRTAVVCSATPARCAGSAPCGVPPP	3014
QY	46	-----	48	QY	90	-----	89
Db	1935	VVVGPRPVLDPDHLRVPLEVACQTAQRGSGVRAQRQSLDVTLVHGGHFRFGHROQG	1994	Db	3015	TCGAVAPPNAPCOXARRRRXRRVSRPGSTWTTACSGSROLARLLHQRDEGGVLTGLQV	3074
QY	49	-----	61	QY	90	-----	89
Db	1995	ARRERFHPGREQPRVGERADQHRDDGFSHLGDDMSRHRRLDLDAASAVPAPEVMS	2054	Db	3075	DAGAAQVQVEGEALAAAREVGVDHAGAPDLDPVAVAVPADEVELPRGQEPQPAELSRRS	3134
		-----		QY	90	-----	93

Db	3135	QPRTDHVRPGBERGADGVHGGGLQPHVVVHEQHPVPGFPQHRVTVGRRQPSRGPODHPRP	3194
Qy	94	:	
Db	3195	PGLVHLGGDLGLGRCHRPVQQHDVFTVQLVEVVPQMPADVLQTEGTDHHTWTPPGGVP	3254
Qy	94	-----	93
Db	3255	VRSLFLVHSHSSMAAFRRRGHSRVS VVIHRTNCFLENVLKNRRLRPHRCKRSRSGVLP	3314
Qy	94	-----	93
Db	3315	LDRXGMLEAGRLIRANDPVLTYRGAVSLQEVDDGIAPWRIPFQERHLFPPEGVGRAAMP	3374
Qy	94	-----	93
Db	3375	TGVRVTFRTDAEGLAFRYAARPAPEMPGPETAHVDRVDRGKPVASLPLVTDREVHTCRV	3434
Qy	94	-----YILAG-----	P 99
Db	3435	GALPGGSDRLVELWLPGLNQFVLHGVELPAGAEVGRDTHAPRWVHYGASESQRGALSP	3494
Qy	100	IQNVSTI-----YL-----	108
Db	3495	TRNWTATVATELGDLTSLAIGAGCYLOPLFATLLRDLPADLITCVMGNVIYGARALNQF	3554
Qy	109	-----WFD-----	111
Db	3555	TYRNLVGLVRIIRERHPSTPLVIASHHYPWHDPLEGDCYLSLTVREQTREVVDLLRA	3614
Qy	112	-----	111
Db	3615	DGDNVHYVHGSLAGPETAHLVVEPRYTDPLHFNQEGHLLAAAFQKVELVPLDVR	3674
Qy	112	-----	111
Db	3675	XLPSRKAULRIEXPADVSGGWSGPSAS PMCPGSPWAKATS PAXRRDSTPSPASIWAIXA	3734
Qy	112	-----	111
Db	3735	LTVRSVTPSSAAISRLEXPMLSRRTWSRSSESKCTWLSQSNWSNRNRSPLSLES	3794
Qy	112	-----	111
Db	3795	QPEYKVPWPYERVVLACPKRSSDASSIKAPKGATTELXREGGDTHATNGSWTLILII	3854
Qy	112	-----	111
Db	3855	LSHLRPGWGLVPPAVVSRARRGVKADKAARARCVHGNHAGRNAATGPPGATAPRRXC	3914
Qy	112	-----	111
Db	3915	TSKNPARRPACPETGRVTCFGRPAMRAPXSTDRCGIYENFPFRATFALATCAFCPARRR	3974
Qy	112	-----FYSTQUR--	118
Db	3975	RPLPFTVQALTRSCRTLGHTSDLRIPSAAXALATRAVLGTRKRSIILLTWYSTPWRLR	4034
Qy	119	-----	118
Db	4035	XSLLAMQVVRSPAMRSMILASRLDKVVGAVL/TWYSIGEECKEVS RAPGRDQAPADS FV	4094
Qy	119	-----	118
Db	4095	PNTRSPAVVDFRSGAGRSWTLGCSSKWGGFRGPAGXPDPFTIMAVKRRLRLGSGXGAPG	4154
Qy	119	-----	118
Db	4155	CNGRIPRCGVPPAAKPRVTGIWPAGKLAGRAHVLADGTRIMGAAADRNFTRARLIL	4214
Qy	119	-----	118
Db	4215	DACESVAMNWRHRHENRCAAADLXPGYGAGRARRNPHLPGGPAHPCIQYAAAPVPGKARP	4274
Qy	119	-----	118
Db	4275	ALSPALCRVPVRRGLGRADHQRRRRPQOLQGIHPLRLARGLHQDAGRGHALRPHLVPD	4334
Qy	119	-----	118
Db	4335	VHRHAQALGRMSAAGSGRAAAAHAGAPVRVAVVGAGGFGVTAVTRALGSAAPVTRAN	4394
Qy	119	-----	118
Db	4395	YEEARRSGPPFDVLNAAACPSRRYWARQHPDDDRRETVDKTRALLRDWMDRFFVQISTISA	4454
Qy	119	-----	118
Db	4455	RTQLDTPYGRNRAEABELCAGHLVVRGLGPMYGDNTKGVLDILEQPVVYAHGESRQSPA	4514
Qy	119	-----	118
Db	4515	PVEWCGGVASHLDAEGLWEVGARTTVSLREIRDANGSVFAGARKDDQFPLVSEPDWP	4574
Qy	119	-----	118
Db	4575	DAADVIGHLRARSALTGESYARPVTRPSSITVAGSATTTPCCPCSTDRRSPACSRGPA	4634
Qy	119	-----	118
Db	4635	ARTCKTFRWSMAARPAVANWCSCATPPTGASCTARATATGPASTAPWRTTCAARSPSP	4694
Qy	119	-----	118
Db	4695	AMSTAPATWSWTSAATAPCWRPTPTGPAWSEWTPPPSSPRPTRRASSXSPTSHTT	4754
Qy	119	-----	118
Db	4755	CSAGAAPRSSPRSCSITTCVPWSSCGRSAACXRTTASGXPSRAICPRCSTPAPTWSAT	4814
Qy	119	-----	118
Db	4815	STWTTTGARSSGWPAPAXRWMPSPPTABASRSCWPAAPLARSSTRWPASAPER	4874
Qy	119	-----	118
Db	4875	RTCPTRSSPGGPRSPVTDWSSPPRGTRGCTPLATAPRRATSSCSTAATRRSCPASP	4934
Qy	119	-----KPAK-----	122
Db	4935	RXTRTSSAATRDARTSRSPRRRPGRLSPTSSSRGSTGTWSPGNATSWPPEAANSFR	4994
Qy	123	-----	122
Db	4995	CPRWKLCDRVAEVTSGTRTVLAVCLLSVLPLLAALLLWHGTTCKDDASRPGSGSATPG	5054
Qy	123	-----	122
Db	5055	DAPHVEAWLLALAVVAVARACGALVGRHLGQPRVVGEMISGIVLGSVLGVPVGH	5114
Qy	123	-----	122
Db	5115	DALFFPAALHSYLDLVAQIGLALFMFLIGMEFGDTHHEGAGRTGAAVIGVCVSFALGCA	5174
Qy	123	---YVYSQY---	128
Db	5175	LGVALTYGAPDGVGLPFTFLGFIAMSVTAPPVLARLLMERGLMQSRAGTYAIVGAATA	5234
Qy	129	-----	128
Db	5235	DLACWLLLAGLVALLRGSPGLGLRTLTALTAVFFGVVWVVRPAURRVLPERRLPDGG	5294
Qy	129	-----	128
Db	5295	VLTLIIPGVLLSNAVATELIGIHLIFGALFGAICPKTAPALADARKLQELVTAVLLPPF	5354
Qy	129	-----	128

QY	129	-----	128	QY	148	-----	147
Db	5355	FASVGLKTDLLQGRGGALWVWAGVALLVAVVVKLAGSAAALMSVERVDALRGVLM	5414	Db	6435	SRVTGSPGWAANCWKPSRSSPRXSTNWSRCRRSGRPGCCWRATTPRSTAPRXS	6494
QY	129	-----	128	QY	148	-----	147
Db	5415	NCRGLTELVLITIGLBELVLPALFTWLVIVTLCATVMTAPLDDLDRABARTAPARTKA	5474	Db	6495	SRCSGSPRCGGHAGSSPPPSATRSVRSPPSPGCRXPXPTGPGXSAVGRCCCARPGV	6554
QY	129	-----	137	QY	148	-----	147
Db	5475	SSVVRXRCGEFCPGNAGAGLDSLXNTDKLTWIVKRWQIRARRVQXAXKQGGVQIA	5534	Db	6555	ARWRWVCSTPRPPSGRAWTRWSPRRRPPXSPVTPARSRSSAAGPTRGWSVGS	6614
QY	138	-----	137	QY	148	-----	147
Db	5535	DLYIGALGVFPPCVSEVAVERGLYPABEAHAELGGVAIAGDVPPPEMALRAAQAVK	5594	Db	6615	PPTWPTARTWPCSTGTCARPTXVPAHRTTRRRRTGRPRXPTARTGRRTCATR	6674
QY	138	-----	137	QY	148	-----	153
Db	5595	RMGSGTEFDLLYASTWHQPDGPPQAYLQRLHVLGGDMLALEIRQGCNGVFSALELAV	5654	Db	6675	CGSPRSPRRPRRTATGRSSSCPTWSTRTTRRWNAAWRTCSSDRRYGATSPRGPSA	6734
QY	138	-----	137	QY	154	-----	153
Db	5655	GYLOADPNRTSALIVAADNYGTPLIDRWMGPGFIGGDALPPWCCXKRGFARLCSVASKG	5714	Db	6735	PPWAPTATAXAWTGRSSRTGTSKSCRPTPGSTARCGVSPGCRERPSAATTSTPTRCWA	6794
QY	138	-----	137	QY	154	-----	153
Db	5715	LPEIESLHRGDEPLFPSPITRGRATDFSARIQQFATRSPAFVAMAEIQDHMDVAERAL	5774	Db	6795	RPAAAGRAATCGCCTARWTTTAAARTAAATPSTAWRSSRPPCWRSRWRPAPRAKSAAPSRT	6854
QY	138	-----	137	QY	154	-----	153
Db	5775	AGAGIGMADVARVSFMYTHAKWSSSAEWRPGGCRPGNGSAGRSATCGASHLLSMEH	5834	Db	6855	XXPTCRXRVGRSRSSARAWCGWPPGCRSTPTPTPTPGSSMPRGRPRRTSPVWRRG	6914
QY	138	-----	137	QY	154	-----	153
Db	5835	IVRTGELAPGDHVLQATAPGLVSSAVLQVLESPOWDEXXIIPATTNGICPYPLIART	5894	Db	6915	RCWTRASTGSNRPTPAWSPGCHPRWAYPRPVSIGASSDCPPVSVYCTLCRSRTPRPGPR	6974
QY	138	-----	137	QY	154	-----	153
Db	5895	PSLRWWSGDAGSPVUSTRPANSCTXRDPATPSARCPTAGKSTGTSVGRSMRRCGR	5954	Db	6975	CWTPXCRSRPPSWASRSSAWSCTSTRSPSTARHRRRRRSRSIPASPTCTPWSRTGR	7034
QY	138	-----	137	QY	154	-----	153
Db	5955	FAPAVSWTTTXRVSPSSASRRARPSXWTRSSAXCWPRGWSTRASTRTPWQGTPV	6014	Db	7035	DARWRACATRWSSSRPRPTTGRAATRTTRTWCKSRVRRRSCVPGCSTRCASRRR	7094
QY	138	-----	137	QY	154	-----	153
Db	6015	CSPGSAPTTMAPDGWKTCTSTHGRSARRCAPCTASLTRSTCAGPASFTPRARPPW	6074	Db	7095	CDSPRPCTFAARWWSRGSTRXXRWWSAAGWRSASAGTCRPTSSGSCPPSATSSITXDNA	7154
QY	138	-----	137	QY	154	-----	159
Db	6075	RCTPPRACGSASAPWRPGASTCWCCHRARRSPWARPARWRPTAGASRPRRDPATPLR	6134	Db	7155	SRNTRARTAMRPERTDSVTHVSRGRRGERAAPAGPETXGKEKHASTIVVGRHRMPSGV	7214
QY	138	-----	137	QY	160	-----	159
Db	6135	RPTATAAARAAACWSSSCXKRTPVATVTCWRCCAAARSTRTGAPTASWPPAGRPRSTXG	6194	Db	7215	PLKRPDLTALTARCGIXRRSYWAGIARGCGLRIPRPNWWSRWPIELFSPMYWIPSRRP	7274
QY	138	-----	137	QY	160	-----	159
Db	6195	APXRPASLPTLTSSRPTAPVPASATPWRSARSPSTDVTPVRSRARSGRXSPTSGIW	6254	Db	7275	EYRRGLSSGRPRRRLPWETARSAWTSPTSSRTPRAGWTPCSPVTSSATTSSCT	7334
QY	138	-----	137	QY	160	-----	159
Db	6255	RGPPGSGSSRRSSPTRTGSRPACWTATPTSTGTGASTGWRPGCRGPSVTRTAGRL	6314	Db	7335	WATTTWTSSTSWRTSAATSPRRSCSPGSPTRAGSAWPNLRTAGXPRSWRSRPTGA	7394
QY	138	-----	144	QY	160	-----	164
Db	6315	CPVSATAAPWPMWSSNRPPRPPAPRRSRPEPCSRCPRRRRTVTPRSPSGSRRALT	6374	Db	7395	TWRSSACTPSARPCTRPWPASGRPGTWSRPTPCSGWPTTGTTRSAPRXSTGTGRTSRAS	7454
QY	145	-----	-VPS-- 147	QY	165	-----	164
Db	6375	WPRSGTPWPTIGSPRWSTGRPSMRPAGTNWPPGCARWPLRNPVPSPPGRCRMPAVPSGC	6434	QY	165	-----	164

Db	7515	TSVPSPSRATAGSRTASSTRSCCVALARSPSVSRTRTXSAGRPRXTPPPGRRPTGWC	7574	Db	8595	AAAIERFRDRGFANAGPVLAPDAIARLKAGAERLITRFTDEGLRSDDYWNFPVEGDERPV	8654
Qy	165	-----ECCCN-----	170	Qy	192	-----PT-----	K 194
Db	7575	WETTAGRSARVEACGGGDBGGLREHARRORCGGTGTPRGHRWSGLHRLPVRARTG	7634	Db	8655	LYRVHLEKQDWAPERDILLHRELAQAAA FVDGPVWPPTAYALVLKEPYAAAEVPHWRDR	8714
Qy	171	-----	170	Qy	195	VNVDSTIY-----	203
Db	7635	PGRRSAGDAGOADLRRPEPGACRRPVHLREARHLRQAAARRGGPRSGQLRGRI	7694	Db	8715	VNVGPRTVCNLSICLDDAGPHNGCLEAVPGSHLLPDDAEVAKVRATGPVVPVPSQGDVV	8774
Qy	171	-----	170	Qy	204	-----	203
Db	7695	ARRPVDRRGAVHTYOCAGRPGTHAGVPGRGHQDRAGLHRRGVRQHRDGLLGRGRADRP	7754	Db	8775	VHDVRLVHGSGPNANGSWERTTIVIEYADPAAPPAPXALRRRAPGRHRTAGQGVFRMRIMF	8834
Qy	171	-----	170	Qy	204	-----	203
Db	7755	QLALRGVQGRHGRPLRPHFWAAGERDAVROQLRALSVPGEGRSVHHHPAAGRPOHPA	7814	Db	8835	TASNWAGHYMCVPLAWALRAAGHEIRVACPPSQERGVAATGLMPVPVLDA PDMMESARL	8894
Qy	171	-----	170	Qy	204	-----	203
Db	7815	VRGWRQPGLGARVRPRAGHPAGRRTAAGAGLPHRGLGRADQSGTDTATTAGRPGRLGP	7874	Db	8895	AFVVOAMVTPQSGPRPLPLHPFTQPMQSLNDFDAGDLRDFWKKSIDAVQRSYDNVASF	8954
Qy	171	-----	170	Qy	204	-----FLG-----	206
Db	7875	GRACTREGTRPALLAVRRPPRRRLRTPCVVRARPRRHRALVRREPGLVGAVERARAVR	7934	Db	8955	GDHWRPDLVYHDI MAVEGALVAALRGVPSVYSPGFIGTETEPGLDLSADPLSCFEKY	9014
Qy	171	-----	170	Qy	207	-----	206
Db	7935	RHGSGRADSTHXGRARPRASPSLTFFVFSRCPDPFGDLMAHCLVTGAGFTGSHVSEA	7994	Db	9015	GVWGRDRIOYAVDPSPDVAVPFLGDALRLPMRYVPYNGAQGADPWQLGPIRGRRVCIVW	9074
Qy	171	-----	170	Qy	207	-----LTA-----	209
Db	7995	LLSRGHRVSLDLSGTAERVEGAHLFTGSVTDVELYDLRLFAEQRFDHVHFHFAFAAE	8054	Db	9075	GNSAGIFGADV PALRHAIDAAVROGAEVVLTAPQEQVEELGALPAGVRLRNCPLLELIL	9134
Qy	171	-----	170	Qy	210	-----ILLRYAQRNC-----	219
Db	8055	AISHSVKSLNYGTVMGSLNINAALRTGVSFFCFASSVAVYGHGETPMRESSIPVPADS	8114	Db	9135	PYCDLLVHHGSA NCYMNIGIVAGIPQLSLALNYDTLTCGRRIDPAGATLTLGLEATAEKV	9194
Qy	171	-----PTTN-----	180	Qy	220	-----	219
Db	8115	YGNAKLTVRELETTMRTQGLPFTAFPMNVYGEWQNMRRDPYRNAVAIFNFQILRGEPI S	8174	Db	9195	EEALRVLFDHRYRAAEKLRDGVERA PSFAVAGLLITRLVADGGLSAKDAEVVDNATE	9254
Qy	181	-----	180	Qy	220	-----	219
Db	8175	VYGDGQVRAFSYVKDIVDVIVRAPETEXAGRAFNVGSSRTNVL ELAQAVRAAGVPS	8234	Db	9255	ARRSAXSGKRRD SGSPGPVPASSAPT CANGSSGVCXPAMWTTCP PPGSRTSTRSPTR	9314
Qy	181	-----	180	Qy	220	-----	219
Db	8235	HPIAHLPARDEVMAVYATATEAREVFGDWADTPLADGLARTAAWAASVGPALRSSFEIE	8294	Db	9315	SGSRCCGPTYXPSRSPSRGRSTTSTWRRRPPWTTIWRCHWRRCGSARPGPTRCGWRWRT	9374
Qy	181	-----VPRW-----	185	Qy	220	-----	219
Db	8295	IGGERVPEWAQCVADRSLGADRXPQAQVSLGSAPDSGGKATVNDLEIRRRVRELEFPWN	8354	Db	9375	APASWPPRRRTGIPSTRSPRATCATSTRSARVRCITTRPNGSPRRXPPTTEARWARTP	9434
Qy	186	TKLYVG-----	191	Qy	220	-----	219
Db	8355	DFVYNGVRYATASTRDYLISQPPQERADAFPAAPFGAKRVLELGALEGADTLAMSGQGV	8414	Db	9435	ALPACSTPTDRGCAATTGGWCPSSTRSPDCRS PSTAAAPRPGCATSRTTRCGDXWPXW	9494
Qy	192	-----	191	Qy	220	-----	219
Db	8415	EILALGREENLRRAEFVMEVHGVNTVNLRLADVESMEFAGLGHFDATLCAGLLYHVQRP	8474	Db	9495	TRPSRARSTARSVPVRXPSVSRRRRSPHSREWSGSPSSARRRRTSPAGAAPTSRPPGRSWA	9554
Qy	192	-----	191	Qy	220	-----	219
Db	8475	WELLADIGSVSDCLYLS THYGSSDGM TLDGTRSTPSAKTTSPRPAASASTCAGWTGR	8534	Db	9555	GSFRCRS PKGSSAPCSGGRAPTRGTWPTTAXTVTRATSGVERHDKPDTGQGADAPERRG	9614
Qy	192	-----	191	Qy	220	-----	219
Db	8535	RXCAPWSRRASPMSCCTSGPRQSATSRRPAARPACPAARTRRKEDASMTTATLDR	8594	Db	9615	TQGLGVGRGRGHAGAARGLSAPXRVSARPCRADRPVGRGAVDPGHRSVDRQRRGRHG	9674
Qy	192	-----	191	Qy	220	-----	219
Db	9675	PRAGAPGRQPAVRGQPVVFGEDPHLQGAQRHLPPHPERLPGRTSGTSGLLRRRIRAQRL	9734	Db	9675	PRAGAPGRQPAVRGQPVVFGEDPHLQGAQRHLPPHPERLPGRTSGTSGLLRRRIRAQRL	9734

QY	220	-----	219	QY	247	-----	250
Db	9735	HPERLRGLPLPARDPAGAAARPPVQLPLGLRRGHQAAPVRVRQELPRGAPGRQIHV	9794	Db	10814	RCAAARTXERTITSWTRPGCAACTARSTVCSPSPRNTXRTTCICARRSGCCGRAAGWC	10873
QY	220	-----	219	QY	251	-----	250
Db	9795	LARHRRGPQARGGRADTGTRVARRPGRRLLPHPAQOCTGCVRYRTLGRQGPVSSVCLG	9854	Db	10874	CPRMASGRSTAGRTCGAGPPTDSRPRPPSGPSPWTERSSPADRAVCCSCCATTAGSTAG	10933
QY	220	-----	223	QY	251	-----	250
Db	9855	TWALSGFWGSRTEPAVEAVRAFDLGVNPFDTAHAYCAGMAEAGLARGLDLLRTRSDI	9914	Db	10934	RRAVRSGCCGPWPMTGTCARASSTTTSTGSSPAKASRARPSRSTWTSXPPASRASRR	10993
QY	224	-----	226	QY	251	-----	250
Db	9915	VISTKGLGLRGDGVVRNSDAGFLRANLTDLSRLSGTYVDVFLVHWPDPVRVPPAETAGA	9974	Db	10994	SRRGTRYDPAPAQASHHVCAAVDPVRLHRRCLRLPLLCGRLDVPQRAVRIRGGAGLOAQR	11053
QY	227	-----	233	QY	251	-----	250
Db	9975	LAGFVEGLARYYGVSNFTVEEMAEFSVVTPQVAPVFNMLDRGIEKQVLPCHAAAGIG	10034	Db	11054	RCRLRAQLPLRTGQSRXGQADRGVQOQPOHAADPPADGARRARVRVRLLSPRGPORDVL	11113
QY	234	-----	233	QY	251	-----	250
Db	10035	VMGWSALAHGVLGALRPGQVFPDDWRAYSPTFQGERFAKLLAAVDRLKEPAERGHVS	10094	Db	11114	PAEQHQAVERHGLRRSGRTYRPRRLVRGVPLLPRLHPVRQHLQPAVRDVLPCQLRLGCQ	11173
QY	234	-----	233	QY	251	-----	250
Db	10095	AQLALAWLAHPSGVIPIVIGAQLPEHLEDSVRAVDLDLDEALRALDELLADAPELDGS	10154	Db	11174	EAPMVVRGDRPVPGRGAVGDAARERPPHAAVLGRRALHAARPLRDAGPADRDQRG	11233
QY	234	-----	233	QY	251	-----	260
Db	10155	DOPPAEREVSXGGHGADPPAQRLVRPVDRRDHTGLRLPLRPGHRCRRRRDPHAHAP	10214	Db	11234	KRRHLQLOPHGAAGEGLRPLPALQERRDRVLRRRRGLRHAPHPAREVGLRRERTPCQ	11293
QY	234	-----	233	QY	261	EQFEKKA	267
Db	10215	GRPAGPALGRAAAVERTHAAPRLPRGHGMAGRRRHPWPGQLRPGQLRQLRPAPPVS	10274	Db	11294	DRGEVAPGRPAABEPVGTAGPAALPRGGRPGQCRVAVARRPVGRQSPGRGEATGDR	11353
QY	234	-----	233	QY	268	-----	270
Db	10275	RPRRRDPRDGRGHRTSGRRRPRGRPGQGGGRGRRGPARGGRAAAPAAPVPRGGLRT	10334	Db	11354	RAGRTRRVVRPCLEQARRPSAGAEVHERRXPVTRDRRRRRSMGSAKALVAEGLRKY	11413
QY	234	-----	233	QY	271	-----	270
Db	10335	RRRPRDARLHHORRPVGAAGRPRRLHAVHPAGLAGGAGPRPGRSRRRAAHAVLRRRS	10394	Db	11414	GSETALDGFVLVPEGTVCALLGPNAGKTTAVRILSTLLKPDGGRALVAGLDVTREAAE	11473
QY	234	-----	235	QY	271	-----	270
Db	10395	RAPVRPPPPRRSAFMDTDTTAPSLASLAVPECHIRETGILLPEHVTAPRRQGVLAVRGLIS	10454	Db	11474	VRNIGVTQYPAVEEILTGRENLEMWGRLYHVGRREAQRTEELLEQFDLTEADKRLK	11533
QY	236	-----	235	QY	271	-----	274
Db	10455	PALESVQEAAGLIDDAMRTRSMHDTWTLPEHPEDAAPVRIEYVMDKSPVMARLAGHP	10514	Db	11534	HYSGMRRRLDLAAGFTVPKVLFLDEPTTGLDPRNRNEVMKVRSMVEGTTVLLTTOY	11593
QY	236	-----	235	QY	275	-----	274
Db	10515	LLLRAMETLVGNFIPTWDSMVEKTTAGAPRLAHRDGMYSDAVATGGRVIDVGIVL	10574	Db	11594	LEADRLADQISVIDAGRGVVEGTPETLKSLVGGDRIVVTLAATEDLSAAAQVARTGT	11653
QY	236	-----	235	QY	275	-----	274
Db	10575	DHAPENDCVMAIPQSNYEDEQVTEADRLNATEWDATGAVPAMRPGDALLHNLTLHG	10634	Db	11654	EPEVDAEARRIDAPVRDRVDLVEIVQLLRGQGHGVVDSLRRPTLDDVFLHATNRAGR	11713
QY	236	-----	246	QY	275	-----	274
Db	10635	APAVGKORRVYIYRPGEVERQLGPHVPEYV-GLKQVLRACLQRAASGEHRDEEPF	10693	Db	11714	KEPVGXWPCWGPAGPSPTAGWSAAPTSTGGAI PCXSSTAXCSRXSRWSGGTCSAAPCP	11773
QY	247	-----	246	QY	275	-----	274
Db	10694	EYRPVEQYRLWDESPALISGLRPPHEEYRWKTXCAACRCWARSPRRRRARAASPPRRSS	10753	Db	11774	WRAAVTTASXRCGCDREXCSVSSPACWPSPARSPAGXPTGSARCRWRSRRAAASP	11833
QY	247	-----	246	QY	275	-----	274
Db	10754	PSASGSPIRAPATGPTPICSTATRWSPSSPGRGAGSTSERAPRRTATCSPAGWRRP	10813	Db	11834	TCSARXSSRSSSSAADCSAGAGTRAPARRRRRRGGCCCGATSSGSGSTWGCCCCPRPP	11893
QY	247	-----	246	QY	275	-----	274

Db 11894 GRMRCPCXRCWRTPTSSRRRRCPAGWAPSRSGTRCPCPPAGSCSAIRASRAAPGSPS 11953
 QY 275 ----- 274
 Db 11954 TRSSWSPGPYCCWCSSCRCECAATGRSADHDHGRRETGDERBTACRPALPDHAPDQGIIR 12013
 QY 275 ----- 274
 Db 12014 TLPGDVAVRGDRRRHPLHRPGGGRRGLRGHAHRRRRHHQHSRPPWRPRQGSRPEADAG 12073
 QY 275 ----- 274
 Db 12074 RAVRHRRPBPGRRLDARRRRQPRRLWRQHRGGPRHRRRCGLVVPAAARXPRGGGGLL 12133
 QY 275 ----- 274
 Db 12134 RRRCAQFGRGAGVQPGPVAAAGVRLRBOQLRDQPAAGFAGRRPGGSGRVRPDGAR 12193
 QY 275 ----- 274
 Db 12194 GGRNGRRRRRRRRGRRTTLPGRRRSHLPLGLPHLPVQRPPLLRGAGGAELPGRRRGRVA 12253
 QY 275 ----- 274
 Db 12254 CAXPAHPEHPRGRCDCRRHRRDOHADRRGRGVRPRGPRPRRGAAPLCCRHDPTAR 12313
 QY 275 -----TP- 276
 Db 12314 SDDLMAALAVITALNQALHDEMARDRCVFGEDVRIGLTQVAKGLHERFGDGRVVDTP 12373
 QY 277 ----YFSYTTSAAL-----NVTNTVYSI- 296
 Db 12374 SEQAFTSLATGAAMAGORPVVEYQIPSLLYLVEQIANQAHKFSMTGGQVEVPVTLVP 12433
 QY 297 ----- 296
 Db 12434 GSGSRGMAQSHDPYSLFAHVGIKTVLPATASDAYGLLLLSAVRDPDPVAVFAPSALMG 12493
 QY 297 ----- 296
 Db 12494 TVEVSGELGPVPLGSARIHRTGEDVTVATGCVHVALAVAEAMADEASIEVVDPRTIY 12553
 QY 297 ----- 296
 Db 12554 FVDWETIRASAEKTGRLWJDDANRMCFGGGEVLATAAEQFDLTARPRVTRPDGAVIPY 12613
 QY 297 ----- 296
 Db 12614 ALVLDQALLPDAQLTDAIRAVLXKRGTTXPXLERERRXIPSVGRGSRXSPSGFTSCAGW 12673
 QY 297 -----TTAARRVSTSTIAYRPD---SSFMKSIMATQLRDLA-- 329
 Db 12674 TGVSAPCWTATPPASTWRGATSTTTRTAVSSMCARPGPRXRSSATTAIWASGRTTISAS 12733
 QY 330 ----- 329
 Db 12734 AAPWSPARTTSCSSTTPGRRPTWCADSPNWRNGRSTASSARCSTATTPTTFRSTSTS 12793
 QY 330 ----- 329
 Db 12794 GRTRPCGWASTPSPATGSPIPPRCCRRGGPHGRWSTPCWVRSSPGPRCCARGCST 12853
 QY 330 -----TWVYTT----- 335
 Db 12854 RCCTPTTTRPTCAAGPAGRAGGWPCTWTWASSTGAAGRPWPARTAGTAGTCGATATTSSP 12913
 QY 336 ----- 335
 Db 12914 TSTGTRRRPPGSGAGSWTSAWAASSAASRRPRPERGRPSRRCAGSPTCPCGPAAGTT 12973
 QY 336 ----- 335

Db 12974 ARCALAAGAARRRRGRPDDEAVEGGAGQHAPAGPGERCLDLRTPTGIRRHVPVRRQPDGRAA 13033
 QY 336 ----- 335
 Db 13034 GTRPRGLPAGCPQCPQPORTPGADRRRAGGRAGGDRVAADRRRRRGPFRQRCHRPGRAAAG 13093
 QY 336 ----- 335
 Db 13094 HRLHQLAPLHPHAGQPRGLYLQPPGAARPLRGRRRRARDPDPORPGALPVRRPGGQGG 13153
 QY 336 ----- 335
 Db 13154 PALHGAGLAAQGGAGGGRVRLRPPAGRARLGGVLRDRHSFVLRDRRADRXGRRRA 13213
 QY 336 ----- 335
 Db 13214 AARPARLRTGAGHPGGRHVRHVRHLVRAGCDGGLRGRRRHSRRHGQRLPGDRDALGR 13273
 QY 336 ----LRYR----- 339
 Db 13274 RSRGINHRLRARSTPDGRAAGIRRGACGGPVVPGCDRRRTVCGAVPQAAGRSOLEVR 13333
 QY 340 ----- 339
 Db 13334 GRGPVDRPGLGTSSVLVQGARLGVGLFGRPHBDQHIGPARLVGEIGARVMRSQAQV 13393
 QY 340 -----Q 340
 Db 13394 HHGLTRGEGDHGPGGRVHRVGHQPDPTVVVDVVRQHGTEVRAGHIVQAAAHHVGLVOR 13453
 QY 341 NPFCBPS-----RNRTAV----- 353
 Db 13454 GFYAPAPAPARVHVALLVPGRCRAVVRLQNGVPVIEVELVRDVQQLGRDRPAPLENR 13513
 QY 354 ----SEFMK-----NTHVLIENET----- 368
 Db 13514 TLDQLGDEFLRAHDLRXPGRPVGILAAAGDEVVARSEVGLGEQAIVDTAQLQRLRVVHE 13573
 QY 369 -----PVT----- 371
 Db 13574 PRQCESVLEGCALLVDHSHFGHPHTAAPFVRHLRXTIRSSPEYIGIFSRIVKNLDS 13633
 QY 372 ----IYG-----TLDMS- LYNETM----- 387
 Db 13634 EXRXIHGRRALPLMRVXYRRDVCKIRERVAMTGHSAIALDVGVVIVYDEPFELALQG 13693
 QY 388 ----FVE----- 390
 Db 13694 TFDLAVTADPSVTLRDFIEDTERFYLDGKTPAERPGIILSSAAADESWVRVRRAWRELAQE 13753
 QY 391 ----- 390
 Db 13754 IFGSTAAVTALAREMPTTVANQPPECADVLAEWGLTAVLEITVLVLDCEGVAKPDPRLIG 13813
 QY 391 ----- 390
 Db 13814 IALARLGRPPGELLVVGNRVDHDVAPARSLGCPAVFVLPEDGYRRPPGVHPDLTRYRAL 13873
 QY 391 ----- 390
 Db 13874 RAVRTVPDDVPRVASLADLAASPLARLGAIXPRCTCCPGTVRXSGRRVSSDPGXRP 13933
 QY 391 ----- 390
 Db 13934 SPLTARARRPPSTGRSRRWSAAGPPDRWRRPTSSWPPASVRSSPNVGTWSRENDCW 13993
 QY 391 ----- 390
 Db 13994 SRCWRCERDTCVSWPAPAGRSTRPRSNRTRGPARGPTFTGTPTSSASNASCAGT 14053
 QY 391 ----- 390
 Db 14054 RMPSPVMSRGXAMSTVPASAFCAAGSCVCCRTGXRPPPGASPCGCSAIRMWSGMTSTTSP 14113

QY	391	-----	390	QY	399	-----KTPP-----	402
Db	14114	GSCSRCAAGSAAAGCPPWXTSARACPLRXAACWRSQVXPAPSPRCAGSGDGPSTGRATG	14173	Db	15194	PGCSPIPRSRPPGSGTWTITASAPARPSTTSAGTSSRTWRTPTCRGSTTRRXARAPVGS	15253
QY	391	-----	390	QY	403	-----TSPS-----	406
Db	14174	WMSGHPAPNSTGGPPSPXPRVSGSAGSVCTTPAGPLPRRARSSDALDQOTQTHPRTE	14233	Db	15254	PAWSRRTRTSPASAPWWTSGAGTGCWCPRCCARTPRCAACCTRRPGSRPPGRCWRRRAS	15313
QY	391	-----	390	QY	407	-----MGFORT-----	412
Db	14234	RCEAPPSYPGSEARHORCDQGLDRPIPSVGLVTRXPVAGSXRWSAILPGASLTGAAG	14293	Db	15314	RTARWRPATSSPPYPRAVTSTCSRASSMAGRTSGPRGFWRTAAGHCRRRTAGSSWSSTCCP	15373
QY	391	-----	390	QY	413	-----	412
Db	14294	IPAAAIKRAFSAASNRTELSSGRASAWACAWARSSCVPTTASWMSASPATSSAKATA	14353	Db	15374	TRCPRTPCRRTTSTTCWSTATGWSARVATSSNSARREARSRKSRRWPEPTCGXSRPC	15433
QY	391	-----	390	QY	413	-----	412
Db	14354	PSTCYSAVARSGVIASSSKAALPACWCYTPTDTRVNSRWXALRXSDSNAPASSVTS	14413	Db	15434	PURLTRSAERPPREFHRAREPHCGIGPNGRPTQOQTLVNSLYCLREQXFTWALPVFPFVT	15493
QY	391	-----	390	QY	413	-----	412
Db	14414	LFTVASTDVAASMPXLSARWTPVDSGSRKQASPARXXPGPKLSATYKSCSTTPGW	14473	Db	15494	GDTHYSRLGDDSVXASAHYGNHWPWHLPRASTHRIRIXSLRNGARAGPAVCPRRSAPAP	15553
QY	391	-----NKTASDSN-----	398	QY	413	-----	412
Db	14474	XTFPAMRSRTVGDSTSEASIASRSACRRSRNRSMWOTKPTLYRSPRGITQYQPSGA	14533	Db	15554	GHPSGERRLAGDEPDRCRAVLPHRDLQPRRPVVRPGILLEAAHRCRGDQSCAADAGG	15613
QY	399	-----	398	QY	413	-----	412
Db	14534	LMTLMPVVGELSGSEPPCOXPQIATLRKISCGRRFPFCLEMTPARPEASMTXSNSSLTSS	14593	Db	15614	HPHHGAQFPDHOQRRPALLSGLHLGQVKGAGVPAERGHALPPAQPURGGQGVRLHGG	15673
QY	399	-----	398	QY	413	-----	412
Db	14594	PVSGSHSAPAAARSPNSNTRLTVPVTSAPCLTAFPRSTWSKSRSTCQERTSGPKIEG	14653	Db	15674	AELPVVRVDRQDPVQPRVTDPCGVVHAQVARGGRREAGPDQAPAGQPGGTRLG	15733
QY	399	-----	398	QY	413	-----	412
Db	14654	QSIQOSSLQORSTPPLTVKPRESTSSAPTRSMVVRHEAIDS PMWXRKGVSRNSFTD	14713	Db	15734	LRRGLRARHDDARPCRAGGPRPRHRDHPPGARSGGFLRARRAGLARPRGGGLPAAAPG	15793
QY	399	-----	398	QY	413	-----	412
Db	14714	MPALLSKEAANDPAGPPITSTFLDMELPTDNGSTSRRIADRNEHPARVIVKIFYSECSR	14773	Db	15794	GGGTALRPEPRSAEARLEARGLVGTGRHGRQRPAPAHROXGRRRRGRSRTGRSVVSL	15853
QY	399	-----	398	QY	413	-----FTD-----	415
Db	14774	DALVESVLLHARLCSIRFGISAELEDRPVKERDGNPRTDRGALESFVSDGFIHLPGAPP	14833	Db	15854	PAGQVGLARABALPSEAVLSDVAGAEFIDWFARCSAAAQCRVTRAPLTELRRWRFDAT	15913
QY	399	-----	398	QY	416	-----PLW-----	418
Db	14834	AELABEACALLWRQLDMDPDDPATWTVREVVLGVRDDVFLRAANTPTLHAAYDQLAGRG	14893	Db	15914	GNLAHESGAFFAVEGLRVRTSYGVPVEMSOPITINOPEIGILGMLVKEVDGVPHCLVQAKI	15973
QY	399	-----	398	QY	419	-----DYLDS-----	423
Db	14894	RWQPADAGRHFVGAVSDEAARGDLRLAHRRLQPLGRGRGRRPGLVRXAGRDPAGLRQD	14953	Db	15974	EPGNHRIQLSPTVOATRSNYTRIHQGRSTRYLEFTFKAGAGRTLVDVLOSEGAWFLRK	16033
QY	399	-----	398	QY	424	-----LLFLDBI-----	430
Db	14954	LPLQVVPRPGAAAPAPVLGHRRGGRHADPRRLPPGRAAAAGAVWRGHLPGRGGGTG	15013	Db	16034	RNRNMVQVAEDVPAAKGYHMLPLTELRLRLRIDGLVNMMDTRTVLACLADTTPASRPVP	16093
QY	399	-----	398	QY	431	-----	430
Db	15014	PAAQVCHGEGRGRLSVPLPGAAGAGQHRHPALHGPAPBAARPAAGARTPARRAVHPGRA	15073	Db	16094	VGDVGAALVLSSTGOGTSLNDMGAVLRWFTEAKSRHQLTAHRTPLRDLPGWRGTPEQITH	16153
QY	399	-----	398	QY	431	-----RNFS-----	434
Db	15074	GSAPGQGHATRGVTRWRRSRPPRPGRTXSGWCSEAEMPPRSWAMRCGCGCPTSRGR	15133	Db	16154	EDGRHFSIIGVTAQVGDREVAEWDQPLLPQGRGVVAVFKVIEGVAHLLVHARFQGLL	16213
QY	399	-----	398	QY	435	-----	434
Db	15134	ANGPPTGWPTSRANPLPYTGCAAWPRSECCASPSPXHRSGNCVAPTARRPSTHW	15193	Db	16214	DGEMMGPTVQCVPENYPQGPFRFLDYVLTASPERLLYSALAEAGGRFYHQSNSYLLVEA	16273
				QY	435	-----LRSPTYVNL-----	443

Db	16274	DDFFVAVDDYCWVTAHQLTRLLRHGYVNVARSLLACLQSWXPHERASRAAGHSRL	16333	Db	17354	QVFARRPIVRVAGHPAHQFMSMRPPQWQDAGTRXEGRGDGAEXAGGGPGRAASRGAG	17413
QY	444	----	443	QY	448	----	447
Db	16334	LRHRPTYPAGRAEGAGDDBRLPERGQPCORRAVRRRRRLRGAAGEGCRRLR	16393	Db	17414	GGRRSTRRRRTGGPTDPXATGPARAGHPGPGPAVGPHRGOAAPGGGGEGRVPDGRPG	17473
QY	444	----	443	QY	448	----	447
Db	16394	LPAHRTAPLDRARPARGKARAGEAADRLPGHRDGGPPRGLPRPDADGEPDLHPALAA	16453	Db	17474	ARRLLVPAAPAPGRAADRRPDGRRPADRRCTGQHRRTHRRGQARAPDHRTRRAARA	17533
QY	444	----	443	QY	448	----	447
Db	16454	RRARDAGRHRXTALPHQVRVPATASQRPVPAGAGRRQPAGRGRVRSARRGEHVPRA	16513	Db	17534	GGGPRGTGRRRGRGADBEARPAACRGAGGAPAAHAVRGACRTGTGAGRGGDPRTR	17593
QY	444	----	443	QY	448	----	447
Db	16514	RPRCGDAADNGBRSRGRRAAVCARRPHGTRVLRLRLPLRVHAVGEGPDHRRP	16573	Db	17594	AGRPRCATAARPGLRPAAPPGLRGGRPGRLRAALRVLRPARRALGPRPRPGRGVP	17653
QY	444	----	443	QY	448	----	447
Db	16574	CLHPAARLPHGSAGTRRRGARARPAGRRPVRRHPRVLRPGRHRRVVGGAQRCHTGPI	16633	Db	17654	AGRGAGQDPCARRHAPPGLTARLPLRLPRPAPDDPRRDGARGQDDPGHRTTRAP	17713
QY	444	----	443	QY	448	----	447
Db	16634	GKSGAGPGTAARQLVRRSRWPMASAFSLVGRGSWAVSCARTXPRPBRMRSPGYR	16693	Db	17714	SRGAEPHGLSGORADQLDPCDREAQCAARDTAARPRPAGRVRLLEGARRRHLLRR	17773
QY	444	----	443	QY	448	----	447
Db	16694	TSPCRAGSSWTSPPRRPGNWPTYSTPSARTWSTPSAATGGSPNGTWRPTARCRPAACW	16753	Db	17774	AAGLPGARRRGHARRRGPLREESDPPLHGLRLVGGALRPRAVPDRIYAHGEBRRGVP	17833
QY	444	----	443	QY	448	-----HRR-----	451
Db	16754	TRCGGPRAVPTSSTAQCWSTARPRPVNRPGPRRRPGRRPTARPNRRARPCWRPPRA	16813	Db	17834	QFGADPPARAGGHRERRRGGIQQVPGGRTGLSAPQAGRPHRTPGAPAHGRVGAQH	17893
QY	444	----	443	QY	452	-----	451
Db	16814	SWRRACGSFTWGPRAPRBSACSAGWPGGSPRRSPGTCSSRSWSCAPTATTSTCATS	16873	Db	17894	GGRGGLPRGARRQLHGDAGGVRASGAVREGAPAAARDXKGGRRERAEGRVLVPGRTG	17953
QY	444	----	443	QY	452	-----	451
Db	16874	RTCEWPPAPVRGWSPSDAARPCRCGSSICSSRXAACPPRSGSFPPPPRABPOTGS	16933	Db	17954	SGTRGAGRGARSGSRPAGRAGVRRGASHSCAGPGGKGSHAGHAGPGDRRRTGHTG	18013
QY	444	----	443	QY	452	-----	451
Db	16934	RSTPGPANSSAGPPSARCTSRXPGCGPTRCAGRQSPSTRSRGRPRCQIPVCRARPP	16993	Db	18014	ADGADAAQAGCTAGRACVRADLGQAAARPPTAARRLLRRARSRRRTARADRRSGPQH	18073
QY	444	-----TPE-----	447	QY	452	-----	451
Db	16994	GGREFDDRAAARAKAPRLDRTTTHVFPVLKVSPPGTPPTPAPADACRPPGATP	17053	Db	18074	AGRLRGDFLRAAAQADRRTPGGRPCPHGPGPLGPRAPPALHGRGRTPGADAGEQVPPV	18133
QY	448	----	447	QY	452	-----	451
Db	17054	GRTGSARVPDSSPRGXSGRCPSPRGRCTALRPGASRRTEADAEPCDQGLLGVRN	17113	Db	18134	RRLRPQRRRVDTCRRRHLGPRPGPHRGGTGTGVSSGRRAYSVAEVKARQAVGL	18193
QY	448	----	447	QY	452	-----	451
Db	17114	EVRPGPVRGARGRRRLRHQADRAAGFLPHLAVEDGTGRGQGPVPEGEPPRLRL	17173	Db	18194	TWRVGPSSGVFLSRMMVMWGORATSTHWPPSLPLKLDLNLHSSAVLIYLPDGLAGSSFP	18253
QY	448	----	447	QY	452	-----	451
Db	17174	RRRLPARRPGEAGADPGRAGGRDPAAAAVGRRRRHGCHCRASPRSGAQRHHDVGEADQG	17233	Db	18254	AESWPVTAASHTCAPGWTTPRGQKPTGCHVSRCPGADRASPGGXRGPRVRRRCW	18313
QY	448	----	447	QY	452	-----	451
Db	17234	DGLRLRPPAAQKRLGGAESAPYPRPHGGAERLGSKARHLRHAGTGGRDQGTDRQAPLP	17293	Db	18314	VETGGRAPVRRTRVGGAGRGGLSAPVRCTAPAPTXXSSSSAGTCGAGSPGVRCDPPS	18373
QY	448	----	447	QY	452	-----	451
Db	17294	GPRSRPALPGVVGHDAARLRVSGTAGAQDRTGFLSRPISAFNRSAWPAPRPREPTXSG	17353	Db	18374	WLPARPRTRAVRABAGPRGRSGSAGRAGRRCTSTGWRWPPAPACTAPTAGTGRSPA	18433
QY	448	----	447	QY	452	-----	451
				Db	18434	CSRPRQPTCPSGRGVRPAPGRVRAAGAAASPVRSRTWSRRRGPLCPGNSAXSPPCGRPA	18493

```
QY 452 ----- 451
Db 18494 SACPPSAWSTGSSAAASWSSCWPNPAPHAPIRTRCPVQNPSTGARTDKSQRCATFVEX 18553
QY 452 ----- 451
Db 18554 TTQTNSTSATSTHATADPSPLPATYPPGHSSTGASPTGTGTTYSSAXASWRTPPSST 18613
QY 452 ----- 451
Db 18614 ASRPAMVSPASANTATSSASRTTAERAYPEPTTTKAAAVCSASARWPTSGAWASER 18673
QY 452 ----- 451
Db 18674 PAKSSGSPSCRALVQLWRPQSGXRTGEAQXVDMPLPTPPRRPHKRTTDXGGPLRWQX 18733
QY 452 ----- 451
Db 18734 CVIPGEDWPPSAXRPPRPSRASSPLSPRPPRAVSVSTRARQTWPMTATXSTRRTTPV 18793
QY 452 ----- 451
Db 18794 RVPRPDPSRSTAPTSSARTRRDXTRSRPPRRRPPRRSPTRRSPPHWSRTACSASTRP 18853
QY 452 ----- 451
Db 18854 RPTPPGAWTARTSPXTARTPTTRPGREXRTSSTPASASPTATSAAGPTATTPS 18913
QY 452 ----- 451
Db 18914 TTTTPRTATATARTWPARSPATPTASPRPRXSASACKTTPARAPPPRSPASTSGPT 18973
QY 452 ----- 451
Db 18974 PSSRSPCTCSAAAPTRPSTRPYATWPPASPSPWRPATSRPTPPRGHPHASPPRSAR 19033
QY 452 ----- 451
Db 19034 RPARTRPATPTTAPSTSSPPVPSRPGCTQATRRTPSPVRRWRPRTWRAPRSTSP 19093
QY 452 ----- 451
Db 19094 TPRPPRSPRXRPPPPASPTPARAPTGSCTSGAGTTTLARASRTPVTTRSATT 19153
QY 452 ----- 451
Db 19154 RSSPRXRSPASPTRRPPSPXRSTSTRTSATSRSSKSPPTARRTRSSRTAPAAVTTST 19213
QY 452 ----- 451
Db 19214 PRTEXTPPRRPTARGNGXATRTTTPGAXTPCRCSSPPXPVLVLAARHVRPTPTA 19273
QY 452 ----- 451
Db 19274 RTTHVRPHAASGKVEARWAGWVRGASPOLRAGGVPSYACCADGARTPTGQSRTAGR 19333
QY 452 -----VN----- 453
Db 19334 GRVGGCPAVTGVNTRHLAOGCRSBDGHPARPRTTQPKATRTTTPHPPQAAAGWTR 19393
QY 454 ----- 453
Db 19394 TPTTAGSPSPSPRPPASPPAYHSPSTRTPVPAARPCPNQEHGTRVHTSAPRTPTAARVP 19453
QY 454 ----- 453
Db 19454 GAAPRAERAGVPATVAIPARRTRGCTGVVPPPLLPQLPRTQRRRRRRPPGRHP 19513
QY 454 ----- 453
Db 19514 LLSAGRAGAGTDRAGADQAGRARQGHGDSRFGALLRTRPVDGVAVHGLHLRRRRRLQ 19573
```

```
QY 454 ----- 453
Db 19574 GRERRGHPLRPOQRLRLGLDRHPGRLRGALLRRARRRGVGRHPPRRRLPGRR 19633
QY 454 ----- 453
Db 19634 RTHPDRGOVLRRARYAMDTHAVPDADVYFIPTRQSWRFGDRREITCLFGATQDDGKL 19693
QY 454 -----LSTNS----- 459
Db 19694 TGLSRSDPTTLDADQVAFLSATNAVDTALYEBPEEYAEEDLAANKAWAKDVHVLGEQIE 19753
QY 460 ----- 459
Db 19754 GCADTAGRAPPSPSPISWRRTPGTGRRRRRRPRTRTRTTTADTSTSTAPRSPR 19813
QY 460 --LWVW 463
Db 19814 ARLWAW 19819
RESULT 14
ABP76680
ID ABP76680 standard; protein; 19938 AA.
XX
XX AC ABP76680;
XX
DT 26-FEB-2003 (first entry)
XX
DE Streptomyces viridochromogenes Avi gene cluster polypeptide frame 4.
XX
KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;
KW medicine; Staphylococcus aureus; biosynthetic gene cluster; enzyme.
OS Streptomyces viridochromogenes.
XX
PN WO200268436-A1.
PD 06-SEP-2002.
XX
PF 24-AUG-2001; 2001WO-EP009815.
XX
PR 25-FEB-2001; 2001DE-01009166.
XX
PA (COMB-) COMBINATURE BIOPHARM AG.
PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;
XX
DR WPI; 2003-018650/01.
DR N-PSDB; ABZ37516.
XX
PT New avilamycin derivatives, useful for treatment of infections, and
PT nucleic acid encoding avilamycin synthesis enzymes.
XX
PS Example 1; Page 68-301; 319pp; German.
XX
CC The invention relates to avilamycin derivatives (I) with antibacterial,
CC virucide, protozoacide and fungicide activity. (I) are useful for
CC treatment of infections (bacterial, viral, protozoal or fungal), in human
CC or veterinary medicine, particularly where caused by Staphylococcus
CC aureus. (I) are more hydrophilic than known avilamycins. The present
CC sequence is that of an avilamycin synthesis enzyme from the Streptomyces
CC viridochromogenes Avilamycin A biosynthetic gene cluster (ABZ37516)
CC ABZ37516)
XX
SQ Sequence 19938 AA;
Query Match 21.7%; Score 531; DB 6; Length 19938;
Best Local Similarity 1.1%; Pred. No. 0.016;
Matches 221; Conservative 83; Mismatches 126; Indels 18858; Gaps 70;
QY 2 GRKEMVR----- 9
|||:|
```

Db 422 GRKQMPRTAPMATPLTGAGRPXGLWEPXGWMVAYGFGAPGAPCGYGDGGYWPACFCP 481
QY 10 ----- 9
Db 482 WGCSDPCGCGPGGGGMDTGTVLLVRAAGPGGNGWTTGTMVSGRGGTGKWGRWXSR 541
QY 10 ----- 9
Db 542 GGGARAHACGLWVGGAARSICLRGSGPRRGVSVLGPAAPLGQEVPRVDAGHCGRT 601
QY 10 -----DVPKMFVLSISFLLVSFINCKVMSXAL----- 37
Db 602 PDTSPSRRTLARGGTRPIGAACVAGDTP-----CSKGLGGGP 641
QY 38 ----- 37
Db 642 YPPRAGLDSLAGGVPGPDVGRPGWPGDRPVPCEDDXDMGSRARTAAPRLCARCS 701
QY 38 ----- 37
Db 702 SRCRSPAVSTCRWPPRRRSPSTWCXCCPNCRRCFTTACTPCRRGRSAGPGGRRCTGR 761
QY 38 ----- 37
Db 762 CGPLRGPRARCRRRRPSPGTRPWCXSCSHRCSRAGPEGGRAAGRTGAGWASPCR 821
QY 38 ----- 37
Db 822 GWXRRWRRTSAPWRPGTGWPRGRRGARRPPRAGSPSTYRRWCWSPSLSSRPXW 881
QY 38 ----- 37
Db 882 TNRGRGRGRSWSRPMPSRRPXWPRXRVTWSRRWSTRCRPPRERXRRPWR 941
QY 38 ----- 37
Db 942 CVRPGPCRRRRGTGWRRPAXRRSGRPSRCRRRPGRCWCPGRSCSARGRLRNPWNR 1001
QY 38 ----- 37
Db 1002 RRCRPCRACRGRPGRCRCRWRRSRRRPPRRSRCCXGRRCXRTFSLPARPSG 1061
QY 38 ----- 37
Db 1062 RCTSCGGSSGPRGPRRAGSGRRCXTCGSPRRRPPGRRRAPPPRRRPSRSRS 1121
QY 38 -----YNRPMRG-- 44
Db 1122 GPSTCARCRRRTSRRGRVPRPNNGRGPXPXGRSCRCRPRRLPRPTGYGPRGRPMRGR 1181
QY 45 -----LVL----- 48
Db 1182 RGRSRGPRWPXRGPRVPAGYDASLPSESSSIGGLVCASTRSGQHVNSLCSASA 1241
QY 49 ----- 48
Db 1242 ERLRPOELHKGPARTRPRLCRAFARPRPTCRPARPGADRGLRRGRGFGVPRFC 1301
QY 49 ----- 48
Db 1302 RGPRGRRRHSGREENHGRAGRGGRWSPTRSRRGVRRRPGPSRRGPSTVTRRSSC 1361
QY 49 ----- 48
Db 1362 GQRGWIIGCVRTGRAGGVRLRRSDECEAXLAFISACTRILYGSTPGYGVWGV 1421
QY 49 ----- 48
Db 1422 RPAXTPRSCRGRTRCPGWWASCRPPARRRARRVPMAKRSSPTRPSRRNRGGGPC 1481
QY 49 ----- 48
Db 1482 CSNSPRRWNTNSTRRARWLPGSRTCRRTSRRCSRSTWCTSGRPSGAGTTTTPSPCAPCS 1541

QY 49 ----- 48
Db 1542 RPATTPARSCASAYRRTTRTSWKPRRRVASNARRSCTASARRXXASSCGSRCTSNSSGR 1601
QY 49 ----- 48
Db 1602 KSTATSSATYASAPNSRGSPSPFYPCAPHTPDWTAPSWRRPITSTWPTPRANVAASG 1661
QY 49 -----KIGVKLQDLKLEILRQ----- 65
Db 1662 FLARTCPSPWRASMRCCGHRPXLRRRRACKXTVWEINEHRTXVVQVELQORGRPVRRG 1721
QY 66 ----- 65
Db 1722 CPLPPHHPRLQEHGRRRPHPGQPHRLAGLHRLHRRIGAATGTHAPNRACSSSTMWRA 1781
QY 66 ----- 65
Db 1782 RARSEMSTAGVDSATALRRAXASNRDLFSSIRTRRSSTPSVQSRRCRTTERTWPMRWA 1841
QY 66 ----- 65
Db 1842 RATAWCSTVGLSCGSORITTEAACMLRPTPPASICASSTAXPGAAAKSSTCCRRAGTIL 1901
QY 66 ----- 65
Db 1902 PEIGNPTRPARGPTLGRPVRSVGAGVSRTAATISSWPGVGVVAALSTAGPRASAVARPSS 1961
QY 66 -----LETTISTKYN-----VSQ----- 79
Db 1962 DFGSSCGPAPGPAASVSASRTTASTSRNAENTTTTFSPSAASSAISRCTFADRSGR 2021
QY 80 ----- 79
Db 2022 AYALRIAMKLPARTASRXASSAVLGSHSVCSWAGSAGSVRTSCWLRXRTGATALRNA 2081
QY 80 -----PVKNLTWN----- 87
Db 2082 LEPATPSCSSMASASSGWRIRTLNRSTFRSMGVFVRNSTRSQCSAHSETAMERRVRF 2141
QY 88 -----TEPP----- 91
Db 2142 FTXWASSTTSMPTSPPGAGKPRSACVVTATPPRPFQSANASCRSGPCGVTRTLRFS 2201
QY 92 -----OYILAGPIQNYSTYVW----- 109
Db 2202 VSRVQLISTLAGQTTMKWLCPSNARCASAMAWIVLPSFISSPRIVRWRARAKRVPKAWX 2261
QY 110 ----- 109
Db 2262 PRSETRRCVSSAWVLTRSATSSGRNPASAGSGARPESASGRPXYSERSSKSTRAATS 2321
QY 110 ----- 109
Db 2322 BFCRRSRSTEAWASGGTFSWSACSVSARISATASARPAARPAARPANSRMRSPRPAAGPASAS 2381
QY 110 ----- 109
Db 2382 SGPSFVSAPPASTARRASGPASTSTCSATTSVAVRSGSLSTSMCTAMLRTASAIRWAA 2441
QY 110 ----- 109
Db 2442 ATWMSAVCPTPSRMRSRVEPASSTWPTVRNPPTSPRRSLPSVRSDSRATGMTSSRR 2501
QY 110 ----- 109
Db 2502 ASCSWIGSRAARTASTARAXSSTTACASXSRPPRATASRLRPIAPPLSPRSRILPPG 2561
QY 110 -----FD 111
Db 2562 RGSRRRLVRGMPGDDPDRPSEHLPALRPLAGGPRSPVEGRDRPAQRSCSILACSRSS 2621

Db 4782 RATAGSQVRSRGRFPARRKSVPRPSRRRAAAAAAPCPRPFGWPFRRTPAGSGPSGAPR 4841
Qy 145 ----- 144
Db 4842 PASPSPRRPSSRGMSLPRPRRTACSTGYVRPSPGSSSSCPTGSRFGWAMKRGVPLVAT 4901
Qy 145 -----VPSM----- 148
Db 4902 GUCTKGWHRXASPAFPVADLSGRSRPTSPASRXVPSAPYSSSGGTSRWEPARISVGAS 4961
Qy 149 ----- 148
Db 4962 SSPVSENRSRSSARPRDHTLXRKILXSXSGITSSSPDQSRASASSARKLASCQPXS 5021
Qy 149 -----TCLS----- 152
Db 5022 VSSGRFVTGNRTGNVPTCVSGLPPALPGELVVGVRGVRGGQAEHVVAHPQDPHPRP 5081
Qy 153 ----- 152
Db 5082 GGRVVGHVQLPPOQGARLFGELRGERTQVDEAVGDEALKCASVSTRVPTILLYRTVAE 5141
Qy 153 ----- 152
Db 5142 LGTNPEADATYSRMKKDXRFHXSVATTLGEVFHNYPCRMFISIRNAPRRETXIISGELHV 5201
Qy 153 ----- 152
Db 5202 QKSARDRWAGGIHGLLAQGRHVREAI GTGDL PALPHRRVDRLLVPHHRRPRRRAGG 5261
Qy 153 ----- 152
Db 5262 RLPRLHQERRAAALGQGLDRLAEDLRSSRPLLGRPRRLPRAPEERGQAGRRHRG 5321
Qy 153 ----- 152
Db 5322 RHQARAVRRPRGWSGVGRHXRAGQVRULHRRLGPRRHLQAPEEPPARDLPOR 5381
Qy 153 ----- 152
Db 5382 RDLGLLAGLAAQFAHRGHQRHQPGLVLGDPARGPVQRRFLRPPGVPGAPGTRR 5441
Qy 153 ----- 152
Db 5442 PGDAGLAGRGI PDRAPHGHERLPAGRTGAGILLRRLRQLPRPLLRRLRRLLRPAA 5501
Qy 153 ----- 152
Db 5502 VHRPPGALORHAGRDVRAHGEQGRHGGGGRGVLRVALPQRLPAAVHPGVRVPATGRQ 5561
Qy 153 -EMLNVSKRND----- 162
Db 5562 GRLLRAGRDDPGARHGRVAGRRRGRLRRTGRGTGRHPRRRHRHARGPRPAGARPGRP 5621
Qy 163 -----TGEQG-----CGN-- 170
Db 5622 AGQLRPPVARGGRERPFDDGGRDARRTQGRSQDGRPRPLRPHRGSPLPDQPDAGNRP 5681
Qy 171 -----FTT----- 173
Db 5682 VQALITALMPGFTTRVWGS PASLRPRMLTRLTVXRVXPSSARRKGGRRRGAHAPS 5741
Qy 174 ----- 173
Db 5742 TAGHPRSGGWRPASRVRRGVRHPASCAGRRTVPAPSRPRSSRSHRTRS PAGRSAKWA 5801
Qy 174 ----- 173
Db 5802 RPSRRSPRAAGRRRPGRAARACTGXRRRRGHSFGPHADRTTPARAPGRPPSASAAHR 5861
Qy 174 ----- 173
Db 5862 TRAGRAGDRBHCSTARPGRHPSAGTARVGGAWRARRRRRSGSGWSPGTAVRPRARR 5921

Qy 174 ----- 173
Db 5922 PPRPSWVPRTPGVPLWTAPPAPRPVAVGRPRGRADVRRGDXGRRRPGSRRRCRPPVPVG 5981
Qy 174 ----- 173
Db 5982 RPSRRPPAGTPGXRPPARAGRQEXAMPSAGIRXNPSPRSPPYGPRAGGPWSORTEARQG 6041
Qy 174 ----- 173
Db 6042 RRGQVGGGDPHRVVRPRRHGTDRAQAGVGPGEVRMDARMPAVLRLQYEHRRAAQTGGC 6101
Qy 174 ----- 173
Db 6102 HVVVDVAHHQLAGGPAEPGQDAAOQPGIGLGDPPFEAVEDHRLQDGGQPPFREYVGALG 6161
Qy 174 ----- 173
Db 6162 RLIRDHRRHFAGEGRDRNRRTNLLGQPPGSPNDPALVGRCRRENARSLRRFPVQI 6221
Qy 174 -----FNP----- 177
Db 6222 ETLGVLDIEAGNTRISGDRKVEGPLKPSQFERLVVDDASDIQRNRTVSSHNSPLPYFT 6281
Qy 178 ----- 177
Db 6282 YVAPRVLHAHQWERRPTAMDSPILPARVXSIHYSPAKADAI FRGGTYRSPQMPDERSRMG 6341
Qy 178 -----FFNV----- 181
Db 6342 AVETVIDEBHRAAFQADGFASLRLVDDTELEWLGVYDRLFSEHADFTTGDYFDIAGRQ 6401
Qy 182 ----- 181
Db 6402 DSDRPARLPQIVRPEKVPPELVESAHFORCRAIAAQLLDIPEDELDYGHAILKPPNYGA 6461
Qy 182 -----PRWNTK----- 187
Db 6462 ATPWHQDEGYMDPRWRRRGLSIWTPLDENATVDSGLHYVPGAHGLPVLPHRHIDDDDR 6521
Qy 188 -----LYVGPTKVNVDST-----IYFLGLT----- 208
Db 6522 GLVTDVDAVPAAGVPVPLASGEAVVHALRAPHYAGP--NLTDQTRRAYVLVFWGAAEVA 6578
Qy 209 -----ALLRY----- 214
Db 6579 DPEPRWLDKDXRRAEABARAVHPLRASAPHQVAPASSURYCSTYRSATVTPHNRTAARR 6638
Qy 215 ----- 214
Db 6639 TSSDAGSPARVRRASSGAKSVYPTTSPTEGTISARQPLPVPTTGVPLTAASSETTVAPGS 6698
Qy 215 -----AQRNC----- 219
Db 6699 HOMPHGPVTAMDMASTACAEASRSRRSPPTSPITGSTVEPYRRYVSSKYSQAGPAST 6758
Qy 220 ----- 219
Db 6759 RRRPOACANAAASSAPLCGETRPMKSRKSSLATWSAADRXRAGSTGIGITGASSPPQWA 6818
Qy 220 ----- 219
Db 6819 RCARELXVQPTGLTGRVVKWCELMKAVPGSGPAGPMTPLWSTWTTSTSAVRSRHSISSGS 6878
Qy 220 -----THSF 223
Db 6879 PAGTTVSPCRPAGLPAGPSRKTSPSSSSPSIRLATHTHMPYPYCGGTGTEIOASLPGTRSG 6938
Qy 224 YLVNAMSRLPRVPKYINGTKLNTMRKLKRQA----- 257
Db 6939 MLTSATFNGLIRTPP-----SPRRARRARAAGPGGTAWGAPCRRASAAPRWSP 6989

QY	258	-----	257
Db	6990	ALRSRPGRRGRRCPXPPRASARRAGRPSPGASRRRXXRGSSGCAACGPGCTCWPSPR	7049
QY	258	-----	257
Db	7050	RRPGACGPGCAGPPASPSPGPGTGRSPSRCAAPRAAARSRAASRRAPNPAHRA	7109
QY	258	-----	257
Db	7110	PASVGPAPSAAPPGDRGSRRRRRRAARPATGPCASTXTRRAPWSRSGTAAGRRCT	7169
QY	258	-----PVK-----	260
Db	7170	PAIPRVKVRAPCPAASRCRGXGTRSPRTFCRGCRCSPPRSPDCRCGXSPRAR	7229
QY	261	-----	260
Db	7230	SRAHARHCRPPSRCSRPPGSRBRCPCARSRTGPASTARWSRXSSRPAADARNSS	7289
QY	261	-----	260
Db	7290	LPFESGPRRAASLQDGTDRVQLGRVQRQLVEHERVGDDRAVRPGHPGPGRIELLRR	7349
QY	261	-----	260
Db	7350	GGQHLAAEPAHPVGVIDDDQPSGLLRRRDLPLVHRVDRARIHHLDRLLVRHGLRQGD	7409
QY	261	-----	260
Db	7410	MHALAGDDGDLARVDPGRPERHRAELAGDLLDRAHQARREDGHRVVRTDRRQAV	7469
QY	261	-----	260
Db	7470	GVRGGGQDLADADCEQAVGMVGLAGHARARTGARHQIGDRDLHPGCHQRELVLVG	7529
QY	261	-----EPEKK-----	266
Db	7530	DLLEHQTEQRGDLVLHDSLPFRRRAGQAGELLRQGGVDDASVAEALVQALGDLGEAD	7589
QY	267	-----	266
Db	7590	PHVLAEDADAVARHLVVQRLVORGDRVKCSHQIITPGRGVFSAXRWCSASAGSARPA	7649
QY	267	-----	266
Db	7650	RANSTASSISVLVSAMAPAVASTRVLWVRSGRARHSPTSASRXFSPTCSSKEWPL	7709
QY	267	-----	266
Db	7710	NRXVRQSRKVPFRRQRSTASAAAASATASASIPSTARAVRPNPAALSTGSPASARSGR	7769
QY	267	-----AKTQSTTTPFS-----	279
Db	7770	LVAXPLFSQNTGSRHRAARLNASSTPWLSPAPSKKATATRPSPRRNDQAAAPAMGA	7829
QY	280	-----YTTGAA-----LNVT	290
Db	7830	PAPTMPLAPXTPLTSAACIEPPRFRPRAVPVPHSSASIRFGSAPLARTPWPRWLVTT	7889
QY	291	NTVYSIT-----	297
Db	7890	SSSVSAAQTAAATASWPMXGMPPTTIPSDSIARQSSNLIRRMVRSXRTARGSALIAG	7949
QY	298	-----TAARV-----	303
Db	7950	FPSSVVGXASGAPTAAARRAPAAVVARRRRPARSRPRXPRDCRTAPGRTPWR	8009
QY	304	-----	303
Db	8010	TAGTFRWCPASASSPARRRCSPASSRAAAHTPPRRPPGAAAAAPGRSRPRXASATAA	8069
QY	304	-----	303
Db	8070	AARAPPAPCRGPASASRPFAVRSGGSXARSASXTCRRSCGPARRPTAPSASGTGRSRP	8129
QY	304	-----	303
Db	8130	GRPRWXPACWXRHTSSPVRTAPASGTRGSHRRPPRTWRCRTTRPPSPXSXGTSRSP	8189
QY	304	-----STST-----	307
Db	8190	ARDCAAASXSCWRGPPSRXRPSPGPPARPSSDRLLPSTGTVGRVQEHVIQRRAAQRHID	8249
QY	308	-----IAYR-----	311
Db	8250	HAVALPAQQLHDLHEVADPVAYRRVDTAGLVHGLGPREAGDLSRGGRQVLGGRQHHD	8309
QY	312	-----PD-----	313
Db	8310	PVAHQLOGLGGLDDPAARVDHRDLVGEPPVGLLQVLRCGEHQGALLDHRDHLDPDLVA	8369
QY	314	-----	313
Db	8370	VARVEAGGLVQBEHLHGHEESGGQVEPAPHTARVVLEPLVGLGQVQLLQOFFRTALRL	8429
QY	314	-----	313
Db	8430	PAHVVEAAPHLQVLAAQODLLDRRVLAGDADVPAHLGLGLAGHIEAGHQGTAAVRLQGA	8489
QY	314	-----	313
Db	8490	QDAYGGGLARAVTEQCAHGALGHDEVEPVQRRLGSLVGLAQPLGNOGSRFRPAHTPPSAI	8549
QY	314	-----SSPMKSIMA-----	322
Db	8550	SSDGSAAFMNALSARRWAAGLLQSRPSHHSASSASTVARRPSSSGLRATDRSSGHCTTL	8609
QY	323	-----	322
Db	8610	ARSASRPSRAKSSRSRSPHRFCGATWSHRFTSVLARRTFATKTSHPFAGMRMRKSTSP	8669
QY	323	-----	322
Db	8670	TPSQDAPITLLKCGKRKTFSGSTVRLXLXTMSTFSAFPVSISSRSIAKWPCKMGKSP	8729
QY	323	-----	322
Db	8730	AKYSRRMRWAFSRVAHSSSSRYGSIITADHFGREFLAQPELTXGAXHITHRRLQVLPNR	8789
QY	323	-----	322
Db	8790	MSRKKGNSNBPFGAVRSARSASAKTSVANRSLMLCRDXDVAPRSRXXQXSHADTRSPASI	8849
QY	323	-----T-----	323
Db	8850	ASRVRMLGLNASVPLTSSGLSCAXRELGAQPTASALSCLKTGSASDSYSSLWYIESST	8909
QY	324	QLR-----	326
Db	8910	QORPXTPSMEAHRWTHGSTMVGLGGWVSRSPASPARAACWRSAGCGPRTARPGPRR	8969
QY	327	-----	326
Db	8970	GLGRRRPGRRRGAGAGAGRPQPGSAAAAAPDRPPASRAPGRGSAAGAADRAVRXAXP	9029
QY	327	-----	326
Db	9030	FGPRXTRPARPGFRVRRRSSATGPARRPRRHAWTAPACRPAAAAAGAPADAGPGPSRV	9089
QY	327	-----	326
Db	9090	PGPWRAHRTSRPAGRTARPRPARPPSGPRRRASRSPAGRRGAGRTARCPLRSRAS	9149
QY	327	-----	326

Db	9150	APTERRAARPARCGGRADRRRSPRRGADRSAPOGSAARXXPPGRWARRPGAPAPGARAPAAA	9209
QY	327	-----	326
Db	9210	CGSAGSPPPVLLVREPQAGDRGLVPPEVLLHRAVLGLLVPVLPGRGPLLQARAQHLLL	9269
QY	327	-----	326
Db	9270	EPHVLGHVRAELTLHLAGPVLVDDPALLAHHRRGAVQCQDVVQGVARPHHGQDGRV	9329
QY	327	-----	326
Db	9330	PLGGVEPVGLGDLVLPVVGLRDPDVTVVLGRVVEVDHVVDATAAGDGGIGVHLAVT	9389
QY	327	-----	326
Db	9390	VPQPGRARGLEHHAIVPGGDEVPRDQGLHGAQQQWARQPQHHRGLVHDVLDADRGRVG	9449
QY	327	-----	326
Db	9450	LVGLQRPDGVVHGFAPGVVDQAGRLLHRLQFGGAQAAHQHTLPPEGGDVLEQAGL	9509
QY	327	-DLA- :	329
Db	9510	PDMAFGDREGGEGGCGIHVECTSPRGERADRGRSATKYHCVRSSAARAARSRPCSAA	9569
QY	330	-----	329
Db	9570	SQGRVVSVPXPRSTSSAHRSSLVVPXSRTAWSPPCSKTTGDCRRSSSAAPRAPTSPP	9629
QY	330	-----	329
Db	9630	STSTLTRSTRASSGTTSSRRTVSTAISRVSARSYGRRRRPLXPGLETTWTXLSRPGVA	9689
QY	330	-----	329
Db	9690	AAAACTVTSGKRRSCSAYCGSGSTAKCRPRGAARMCMVRISSPPAPMSRQPESGAQTR	9749
QY	330	-----TWVYTL- :	336
Db	9750	SMISTVNSYSQSLSWGTSISVATSDRPLFARGRLVPAAVQLGRVQQQFVQGAQLGLVEVE	9809
QY	337	-----	336
Db	9810	VHGPDAVLQVLKLGADDDSRGVQHPGKRELGDVTPAFGGELLQPVHGREQLGEAFA	9869
QY	337	-----RYQNPF- :	343
Db	9870	LEGRGVGAPVVGREHLAGTERRSRQHPVRERRPPHDADPGSGAVRQHLLLDAAVEHYERHL	9929
QY	344	-----	343
Db	9930	GDURGGDGGFHFIDGEVDAHVPOBPFLHEPRQTRRLGRDRDPRVGPVDQEHVHVGP	9989
QY	344	-----	343
Db	9990	QRAQAVQVQGAQBTGVAVADHTVAPQLQAALRDDDVGATGAQVQAQVQAQSLRGHARAV	10049
QY	344	-----	343
Db	10050	GVRGVEVDAQVERAPDRLDRLGLPAAPPAQRPQAQDGGDILGALPSQRAVPHAASCSL	10109
QY	344	-----CEP- :	346
Db	10110	RRVRKQTPPWSGCDPTGTVSPASASLRTTSSVPSQDNPLTATGCAPXLLTYSWNCGLV	10169
QY	347	-----GRNR- :	350
Db	10170	TSPKSRKXLRNRPGRSTRRIISRRXRTSKAFRWMKTLSPANPASKEASGSPARKSFRMMW	10229
QY	351	-----	350
Db	10230	XVTLGRPAQVAVFVSRRNDRLTSAHSRLPTRGASSRAMTPSPLPISNRSVSGVRDSAAANW	10289
QY	351	-----	350
Db	10290	SIRSARSRRYSRXXYPASRTCMPGSSSHSQTTLSSSTLRGIRTPLPAGVWVWPLHSXC	10349
QY	351	-----	350
Db	10350	RSNRRRRGPCPAGRCSPTRAGCAXALRRAAPRASSPASSWRSRCRGSAGRARPPPARGT	10409
QY	351	-----	350
Db	10410	RSAAPLPRVRRSPRSDGRSSHRRCXCPGPGRTSPGSPVPAPCPRRSTATSGRRCRX	10469
QY	351	-----	350
Db	10470	WXAARRAPGRXESAPPARRRGASAVRRSXTGGQCRCPRPACARRRRSAPRXTVWPRTTP	10529
QY	351	-----	350
Db	10530	NAGRCGRCPSSSRTAGARRDPRRSPRRPRSGRRAPPPAAARSRSRSPSPAASPAAP	10589
QY	351	-----	350
Db	10590	DSPGGRPATPGRRRGPAFRRTTARSRRPAAAXTAPRRRARRGSRSCRTWTCPPRRSPTDP	10649
QY	351	-----	350
Db	10650	ARXAVTGGSRXSRHPXRRPAPVPSSARLSRAPPSLGGVVHDLGDVLRQPAVRHEAGQQ	10709
QY	351	-----TAVSBFMK- :	358
Db	10710	PGDRGRRCPDLDAVPOFLRRPAAVVEQHPAQRLDLLGRGLKAGQGGAGGIDTAAAD	10769
QY	359	-----	358
Db	10770	QRVVVEGQLNRPHDAVHVAVGAAVHQVAVREDELQRAVAQHPDAGRQSRQLHLHF	10829
QY	359	-----	358
Db	10830	LRRREHHLRALPHRRVDGVPQRRYVGAEDARGVAPDDAHASAPDRPQLPRVPALGAVVG	10889
QY	359	-----	358
Db	10890	HVAHQPOQVQAQRHHRVRXGVHRVLDPVAAPLDAVLLEARQVRRDQVEPRGLHGADE	10949
QY	359	-----	358
Db	10950	PGRHIDARHAPERGQCALDGHVVHHQVGPVVAEAHRVVIATLALHRVDRLLPVQAQVAG	11009
QY	359	-----	358
Db	11010	VEVVERLHRLPGRVQRQRPRTGLRRCVHGLHIERQPGAFHVRVVEHRDRHQAGRGDPA	11069
QY	359	-----NTHVLIR- :	365
Db	11070	FLRRRARHDLVSRGAQRPGQHHAHVMRSPVRGGEHNPBKDSLTCRSMASWRSAAQRS	11129
QY	366	-----	365
Db	11130	GRRRCRVRGVLDGPPPTAVGVARTVDEPDVVHHHVALADRDRNHGSGRAHLRHLGVV	11189
QY	366	-----	365
Db	11190	GEQVRSRNFLOASVVRSGVVQADAQIADRADVDVAVTPVHLRGPVRLLEDERVCGGHD	11249
QY	366	-----	365
Db	11250	RAVHERGGBELQFLAVQQVPRGPVLLQVHVAVEDGTLLIADLGEVPPVVVGAQALVGEAG	11309
QY	366	-----	365
Db	11310	DEFFGARLETGDRVRGENRARVREAVPSEPLDRRGVQCRRGHGCVLLSPRPGSVTSGAR	11369

QY	366	----	365
Db	11370	RPCRRRCRRPLPWPAPAAAPRHRXSPAPCRASATAGPASARRXGRGFGARGLRGRG	11429
QY	366	----	365
Db	11430	GPRTEVGHSVRAAPVVGGEVAVGDGADVGEQLPGPLHVVEPCAEARGVEVABPGELHR	11489
QY	366	----	365
Db	11490	LHVGEPEFHVGDVHLHDELGAQVLFPAALQGEDLDAGLSHRQGVREALQALQDALGP	11549
QY	366	----	365
Db	11550	REEREVRGFLGLRLAQOVVARGGGVDPVHVHKVHPRLEFTYTPNLKIVHCCLTSTV	11609
QY	366	----	365
Db	11610	RCASKRDGLGWSPVGNAPQVGHITLRLRDALSADLDLEGAQFGGADGGGPRGGPGEV	11669
QY	366	----	365
Db	11670	RQGVRFVAEDLAGLLGGVGDHHLVPGQVDRGVAGDTGRRTHGLQLQHGVGAGRADV	11729
QY	366	----	365
Db	11730	EGPAPGLLLRGAHDDVHDVLDVAERPHLTAVAVDRDLTAODLVEENGDRDIAIGVAHVL	11789
QY	366	----	365
Db	11790	PFAVVHVHTECGEQALGPHRRFQALDREFRVAVTVGRYDRDLRPHRGLAVPVHRDGR	11849
QY	366	----	365
Db	11850	ETEERDAGSORRDEVHAAHHVCAVVEALHVRTDLRRERREVEDVESLLGEEPVDQLH	11909
QY	366	----	365
Db	11910	IHRPGEQVREALDGLGAAAEVVEDRYVPVAAEQSLRHMRADAGSAGHTVVRHQISWK	11969
QY	366	----	365
Db	11970	IKWTTNKEFGQRRGSPRAGSPLVMGAVGTARAMPDSARSFNGSHQSFSAVQRTVSAR	12029
QY	366	----NET----	368
Db	12030	PCSNDRTGAXPSAPRASHSEXRRSCPPFRSGTRSTRSQPAPRASSSCCVSSRLVSSAEP	12089
QY	369	----	368
Db	12090	MXTCGSPRSATSRMPRTWSDTCTQSRTLPPSPYSGMLPSSRVVNSGTTSPGNKXGP	12149
QY	369	-----PYTIY-----	373
Db	12150	XLLPHRVTLTGSWVRAXARATWSPALDAAKGELGAIGASSOEVPVSMPLFYTSSVETCT	12209
QY	374	-----GTLDMSSL-----	381
Db	12210	ILAVPASRHTCMSAWTPTCLVRMNGAASAIIDRSTCDSAAKLTTTSWPGTTSSASSLASQMS	12269
QY	382	----	381
Db	12270	PCTKYRPAATGSFARFPAAXVLSSTVTRAGSPSTSSRTYWPMPKPAAPPMTTRCPRSA	12329
QY	382	----	381
Db	12330	PALPSRMFPXAAFTVASPASAGLFNRRPGCGLPAPAGARSTGGRRSPRSPGRSPSPR	12389
QY	382	----	381
Db	12390	HGRLPRSKHAARSSTGSCARPSCPRWEXRDRGRSPAPPCPCGRXRRDRRCGSRSSPP	12449
QY	382	----	381

Db	12450	PPRLGPPVVRWPRPRRRRCGSPCTPAGAGSSGORRXXRRSPSTGSRSPWRPRARGRPAT	12509
QY	382	----	381
Db	12510	AGRXXSPARSARTAGCEPRPRAPRSYXRTTERRSPGRSWGPGASPRPSPGSRQVTRPT	12569
QY	382	----	381
Db	12570	CAGRRPGRAXPARRACKGSGPPSRRCRPRRSCPGTXRSRRRRRGSHGTARRRRPARAGR	12629
QY	382	----	381
Db	12630	PACRXRPGRTPRRLPRKPGSPRSPROSRRPPIFRPPXWNVHRREQVFWPGVWVWR	12689
QY	382	----	381
Db	12690	YAXSAASQPRPIIXAPSVYPTPCGQRQIRSPQNSRWHMSPNNDRGTGVFRSLGF	12749
QY	382	----	381
Db	12750	RAGWRGFTPPSPADMSYGVSAFRTGHRPPSGVGVSRGDRRCRXWMAASRRCWPARG	12809
QY	382	----	381
Db	12810	RGASPAGGPPSSPSRAPAPAGGERAGPWGRVAPSPARSARAPRAPGNAAPATDPT	12869
QY	382	----	381
Db	12870	TPRPRRRPARCPCPARPAARPGSAGRAGSPGVPPRYPXPGARCERRCDRARPRSPPP	12929
QY	382	----	381
Db	12930	PVACRRXSRRRARPCCAGGPRRRPARSTSRPAAGPRTRRRARAPERAVHRDRTVAR	12989
QY	382	----	381
Db	12990	RSNRNRSVRPPRPAAGRPGRCRVRAGARPGPAATAPDDRGPAPSPGRHGRARRGSG	13049
QY	382	----	381
Db	13050	RRRPRSGRPAAPPRGRPGSNADPPSAPGGSSSCPGGEGALRPRRRPGTYPAPARRPG	13109
QY	382	----	381
Db	13110	RSRPRXGRCPGTGGCRRPACCATAAGRCPRRRRACPAACGSRRRGRARPLAPRRWNATA	13169
QY	382	----	381
Db	13170	GGAAARGAAGLRGSRGAGAPTSRSGRRRPRRRRSGRAGCAVASVRRTRPRRVR	13229
QY	382	----	381
Db	13230	PASRGSSARPGAGTARTFGGRPRPRXSRRAAPGCGAGSPSTRRRSWPAGRAPPS	13289
QY	382	----	381
Db	13290	CRAACAVRWSARRRPRAAAGRAGGCGCGRPRRRRPAVRPAPRRSSGRRSPSRGHRH	13349
QY	382	----	381
Db	13350	SRWTSSTRWPRPRRCHCPRAARRPRRDXPSPSRHAPRPAATAATYGRSAGPRRSAP	13409
QY	382	----	381
Db	13410	RXRRRRSHRPSGRXRRRARRPVTTARRARAASRRSWGRRRRCRCPRPPAASAGARTRS	13469
QY	382	----	381
Db	13470	PXYTPAVRRSPRRRPARLPVAVGSPRPTGTRDPTPRDRGRHAPARPRXRDRARVPEWP	13529
QY	382	----	381

QY	399	-----	398
Db	15750	QHRVTSREFSGRDACYPRFRGRNGHATADAAVTPRSATTEPSQFADRDHREVRSA	15809
QY	399	-----KTT-----	401
Db	15810	RRPCKPAPFARTTQCPRATGTYTKVHYCWXTSVGYKGVSRCLVPPRRSHLSALPPQG	15869
QY	402	-----	401
Db	15870	VPGENRPHDLIQGRQDHRTHRRXAHHLHQALPOPPGGRVPRQDDGALPGAQPHRP	15929
QY	402	-----PTSPS-----	406
Db	15930	RRQGLRAGNPEIRGVAPOQSAGTQLSHCHGRXAPSPSWTKRACQCSLSKRVFIYP	15989
QY	407	-----	406
Db	15990	ATIRXLGAHGRPTKAGDAPRFRAGRPFGCRILRGASAPGRGARGARTSRVSGVIAVH	16049
QY	407	-----MG-----	408
Db	16050	TPGTGRLIRLAHASGAPHDSRRHQAPTRPXMXXDDQDEXCPPIRMGGIALTSXFS	16109
QY	409	-----	408
Db	16110	GGALRLDTRCVGXPLRACKDHTLIRGPDFAIFLRLFXFWRRSISPRGPVLAGRAQPG	16169
QY	409	-----	408
Db	16170	TFTLRPAAGLPAPBPRLLQPDORRAARGHRAHQSGYSGDTEARGRIPSSGRTGCLRP	16229
QY	409	-----FORTFIDP-----	416
Db	16230	RASRTHRCRGIRPSTYVGFPLDPCLAGRVESGPHQVTHQLHELPLESGRQVVPLL	16289
QY	417	-----	416
Db	16290	VEQIRIGVRLDVQVGFGAGARAVDVVDLVAVRAQQIDHLTGLLPLDROGQVAVSQ	16349
QY	417	-----	416
Db	16350	RVVPGRVVGRDDQDRGVSLPDDPDQDQVRPVGELVQGTDAIDVHADHARDQIGQVA	16409
QY	417	-----	416
Db	16410	EQGREERLEVAARPDRQGVQVPEFGDGRGVPVGGGRQPSALRLGPGVVDPAGRGVNP	16469
QY	417	-----	416
Db	16470	AHLGACRXLHAVQDELVOAGQPQLHKAVAAARQCPDPAVDLPVGHQGEAGDRLAVHSYV	16529
QY	417	-----	416
Db	16530	DVRRLGARHLGGRTGRVPECEPLGVGPESDPHRRHRGASHAALGKBQVFFLKRYSPGR	16589
QY	417	-----	416
Db	16590	YSTVDLLEGDGAIGQHRVIRPNEAPCFQHPSPVQSFVXADASRLTAPMRSEPPILXHS	16649
QY	417	-----	416
Db	16650	XKAVRTMDDXHSRVPPPAERSHTGVTVDKKTADRNTPGRRPMVAVCAPLENVRHL	16709
QY	417	-----	416
Db	16710	RHNDQNGDEYVLLDRPVTPEABKVATQVNEAGTMRILGATRGLSASRNTVLEWAD	16769
QY	417	-----	416
Db	16770	RHVLFDVDDVRLASAVDAVRAAFRAGAHVVGARLPPPRELRLPWFLLSSGQFHLVGWHR	16829
QY	417	-----	416

Db	16830	DRGDIKIGACMGVDADFARRQGLTFDLDLSRTGVNLQSGEDTSFTIALMKEAGARELPAA	16889
QY	417	-----	416
Db	16890	RABGRPRCPRPAHPPLPPAPGLLACGVRRRDCTSRRRDSARSXPGTAPRRSRAAHCCS	16949
QY	417	-----	416
Db	16950	ACTGPPRSASATDSCCASCASCLLPDGTGTDPGVGGDAFLQDRGRGPAQIGGDTSD	17009
QY	417	-----	416
Db	17010	VQAVLIDLARPAAGRETGRQVAPGGGGVVEDLQHRAPLARVCRRRADGAVRRDERRGD	17069
QY	417	-----	416
Db	17070	VADVHVVTGLRTVAEHRGHSALGGQPAVDGDDDLTARALPGFHVHVAEPQHAVLGSVQPV	17129
QY	417	-----	416
Db	17130	EQADVLLRAVLGRAVGRPRPGQRVLVRRDRNAPFVERPAGGSEHDPDPGVRRRFQHVDR	17189
QY	417	-----	416
Db	17190	LDVDPVDRGCGHRDPDVLGREMADDVGARRPHGLVQGGIGHVDTDDRTWVQTPRVT	17249
QY	417	-----	416
Db	17250	GGQVVOHGDLPGLQBGVDHVRADAEARPSGHQQTCSGXEPREHGHGPPFLATTASECGR	17309
QY	417	-----	416
Db	17310	RSGHIRENLRLDGGDRADRPASPGTTDSPGAGDGLCVRRQLGASAGAAAALXSLQCGDG	17369
QY	417	-----	416
Db	17370	GSEVLGSGQVSAYRHAVERIDSSDLACGVVLHSAGYPAPFPVRLATEIFORALARLPDG	17429
QY	417	-----LWD-----	419
Db	17430	PVTLMDPCCGSVLLTVLGLLHRRSLRQVITASDVPAPLEAAKNLALLSPAGITARELE	17489
QY	420	-----YLS-----	423
Db	17490	RREQSERFGKPSYLEAAQAARLRLRTAEGGALPCAIRTADVDPRAISAVLAGSAPDV	17549
QY	424	-----	423
Db	17550	VLADLPYGSARTGKGRCSRWRALLRSLASALPAHAVIAVTDRSRKIPVAPVKAERLK	17609
QY	424	-----	423
Db	17610	IGTRSAVLVRAADVLEAGPAAVPAEFPCQCFQCPCFPSCFSLDSQRMWLQVCPWE	17669
QY	424	-----	423
Db	17670	CGMAEARVLVAGNLHWQAGFSQTVAAYVEAAGEAGCEVRVSGALSRMDEQIPRHLPEPDP	17729
QY	424	-----LLFLDEIRNF-----	433
Db	17730	VRWGTHLVFMFEARQVLSDEQIELACSFPDRRLIIDFKRPLGRCHRRRQLRRHPLDRE	17789
QY	434	-----	433
Db	17790	LATALLRAORPRPAAQVDGPLARRGRVLPFRNAPAGTASTGARSAGLRAAVHRQQLGA	17849
QY	434	-----	433
Db	17850	VEAVHRGGRGTQTPAAADSGLRVAVRGDLPRLRGHLGRSRNGXSGVEVQPPVPGHV	17909
QY	434	-----	433

```
Db 17910 VTEMGRAVITTVLVGSLVADTGLFTARMLETLAGSLFVLPPAEFMSSVYEGDVEALT 17969
QY 434 -----S 434
Db 17970 GSDPAATLGRLAGDERHAEIVGRIQDRARARYNPRLRELLRFFRXSAAVPLRRGRS 18029
QY 435 LRSP-----TYVN-----442
Db 18030 RRSRPSASSRSPSGCXPCSRRTIYARRCXKARLSHCBSAAGSAGXSRSTHRSRARPYRP 18089
QY 443 -----442
Db 18090 GRSACRRPGPPPPSPRPVTPACRRAGRTASRCSGTGISGRRRRRAPPARSARPCHP 18149
QY 443 -----442
Db 18150 PSEXCCRSPDWSAGRTVPGSPGSGRGRRRRSPRRRPRAGRSPSPAPSGPGPS 18209
QY 443 -----442
Db 18210 RSAPSAPYEEAPAPTSTAPAVRRRGACVRSRASPARRSGGLPSPGTHRRACACCGRAV 18269
QY 443 -----442
Db 18270 PATAGNERCVXSABPPASALHVLIPFEGAYLPLLALTKLRAPPXGRPGCRDVARDOC 18329
QY 443 -----LTTP-----446
Db 18330 ALRPARSAVRFPAGGRAARPLAPRPDTPLRPGSRGPRPVARRPPGALLVAPHLWR 18389
QY 447 -----EHRRV-----452
Db 18390 HGHPAHDDPHRDPRAARRRRRLIRPRAAGHRXRNPNYDHRAAYPQLRPADHRH 18449
QY 453 -----452
Db 18450 GLRRGSRPTLLPRDRPAGRXRPORQMGAGPXPNAPAHFTEILINDLADHAQVILHSRF 18509
QY 453 -----452
Db 18510 WRVADRTYGISGTGLDWELDWSAPWEHVESHTWSLLESEAPVGGSI FVAPTWDRT 18569
QY 453 -----452
Db 18570 DLYPERMPFQHSSESGYGVIERVYVTSKSDRPRCDLKSQVLRPGWPHAERFTHL 18629
QY 453 -----452
Db 18630 IDDAATITHAEWEALLGYEWSRLTAANWLGVRATFRERIGDLPPASEFCFSGGAHCF 18689
QY 453 -----452
Db 18690 TLARFGTHADAEILTGTRQMPIPNNSALAWTGPYCPIGRXXSNSQSAGGTKSQGPD 18749
QY 453 -----452
Db 18750 SDESDESDURIQDDVDLRLNVAEKLQSHLRKRVGAHPSTIPAREDLHGSQXWY 18809
QY 453 -----NLSTNSL-----460
Db 18810 RAAARITISLVTRYTRCSSLIRRDHAPGKWYFSGGLPIPVKGSRLTSSMLLWASIL 18869
QY 461 -----460
Db 18870 RSFLSQDVXSSQPSKSTSLTRHPNORPNPDRSAPATHXRGQRSTPXAVDGRSPES 18929
QY 461 -----460
Db 18930 AACTSPRAALHNPRPRVRCRAWRRSRWDRCDXPSXAERPCGFARGLSTYDFPFP 18989
QY 461 -----460
Db 18990 VVNNVPLLTSSGRAKTDGKSAPPAPSGHQKQGAAPWRSRPLLTCAFESSVVVELRLHV 19049
QY 461 -----460
Db 19050 EAELDHVPVLHHLVIALHTLPLRAGLGDGAGLHQVGEADDLRLDEALLEVGVDDAGGLG 19109
QY 461 -----460
Db 19110 GGGALLDGPAGFLRAGRVGLQAGVEADAGEGGERAGLVADGLEEFEGVLVVELGEVG 19169
QY 461 -----460
Db 19170 LQLGVEEDGLGGDQGALLVLEVLVDQLVLVDVEDVEEGLGGQVQVQVQSLVALGALDDA 19229
QY 461 -----460
Db 19230 VGBEGVAPIQLDLLGLLDGLHLRREVLLDPRVLLAEQQLDLRLREVREDQLGVDRLDVVLG 19289
QY 461 -----460
Db 19290 RDLAVDMDDVLVLEGADDLADRVGLTDVREELVAQALPLGGAHADGADVDEVDRREDAL 19349
QY 461 -----460
Db 19350 GGBDLQLGEACVGDADHTDVGIDRGERVVGRQHVVLGQVGOGLADIRQTDDSDRERH 19409
QY 461 -----460
Db 19410 VATSRYXEDPRAGWAGARSFFDLRCXCRRTGRTSTSLRAGPGPGAVSNTWPSAPRVX 19469
QY 461 -----460
Db 19470 WAIRHISRPLVQWSHTGRDLPGTKEDTLRRRCRRRRRGVGVSLAAGAAVRGGVKTVC 19529
QY 461 -----460
Db 19530 SEAGADSFQAGAVRRAATAVSRAGAVRAGTAGPQPGPPGSRPGPPAPPSPSAALVR 19589
QY 461 -----460
Db 19590 PVRPRRPRYAGPRPRPRPCAGFTTQXPGWAAACSAAPXCSCSATSTRCCSGRRSRC 19649
QY 461 -----WWW 463
Db 19650 TACCPCASCXRRCCGACATWPRRWY 19677
QY 461 -----460
RESULT 15
ABB58144
ID ABB58144 standard; protein; 7107 AA.
XX
AC ABB58144;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 1224.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
PR 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
```

XX WPI; 2001-656860/75.
DR N-PSDB; ABL02247.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 1224; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from wipo.at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 7107 AA;
Query Match 21.5%; Score 526; DB 4; Length 7107;
Best Local Similarity 3.1%; Pred. No. 0.0017;
Matches 217; Conservative 104; Mismatches 135; Indels 6550; Gaps 71;
QY 2 GRKEMV----- 8
DB 25 GRKRLVICKVTDAGQFKCTTNAADTTESEIIINYQNRNKKLKOTEAREKILIDIE 84
QY 9 ----- 8
DB 85 LQDTAPCDWKFNGEPIVPSIESIEIKNMGGKHQILSFSLDMNGEITCESGOLSKCK 144
QY 9 -----RVPKMF-----VLISIF----- 22
DB 145 LSIRKGESRNIDCPDKFGSIPAPVLLVFPFKVSGTKPTVPBAKLFKDGKPLPVKDVEV 204
QY 23 ----- 22
DB 205 AVTDDKVTFKKPSRDLSPYQIKISNGGEBTDKVQIICQDVPQPPQDVIDTVYQTS 264
QY 23 LLVSF----- 27
DB 265 CVVSPNPPSDGGTPTTKVYIERQDLKXGHWESVAEVLPSBCLKKIDDLIPKKQYRPR 324
QY 28 ---INC----- 30
DB 325 IRAVNAIGOSDPATFRNTILAKDPWDEPGKPAVDLTDWDKHADLKWEAPETDGGDPIT 384
QY 31 -----KV 32
DB 385 AYIVEYKEKFSNDWVGSGVDGARTATVDGLKEGQQYEFRRVAVNRAGPGEPSTKSI 444
QY 33 MSKALYNRP---RGL----- 45
DB 445 IAKCRPVKPTVGEGLKNVTVKKGQTRFDIKYDGEPEPAATVWKGTDNLKFDNQICLD 504
QY 46 -----VLKIGYKL----- 55
DB 505 QLERNSSITTKSVKRDYTKYKLVLSNSGCTIBSEAQVVLDRPLPPGPFPEEIRASH 564
QY 56 ----- 55
DB 565 IKMKWRPDDGGCEISGVALERMDEETGRWIPAGEVGNETSDFPKGLTPNKKYFRVK 624
QY 56 ----- 55
DB 625 AINKEGESEPLETFDAIVARNPYDPPSPQPVDDVDNKSLLKWKRPSPDGRPIITHY 684
QY 56 ----- 55

DB 685 IVEIKDKFAPSWEVAKTDDPNPECNVEGLKEKNVYQFRVAVNKAQSPSPQTDNHL 744
QY 56 -----DQLKLE-- 61
DB 745 KHKNLKPQIDRSFTRKRTIKSGRTHKWSVDVLGEPPELHWSMRDDIPLTNGDRIKENV 804
QY 62 -----ILR----- 64
DB 805 DYHTDFSITNVLKDSGFYTLKAENRNGIDRETVELVLGKSPSPKGLAVSDVTASGCK 864
QY 65 ----- 64
DB 865 LQWKPEDDGGVPIKEYVVEKMDATATGKWRVGRSGEKEPPSPFDVTGLSLGSEYMRVS 924
QY 65 ----- 64
DB 925 AVNEEGESPLTLTLGVAVKADPDEPNKPTPEVTDYDQNSISLKWAAPNNDGAPIQKY 984
QY 65 -----QLETTIS-----T 72
DB 985 IIEKKNKTEWEKALEIFPDQLEATVAGLQEGYQFRVIAVNKAGLSPSPDASVQIV 1044
QY 73 KY-----NVSKOP-----VK 82
DB 1045 KYKLLPRIDRSNLKLLIRAGKPIRYDVNVRGEPAPVITWYQNDKELKPEELPSSSEIK 1104
QY 83 NLTMNTEFP-----QYILA----- 97
DB 1105 NIPYNTKISIIETVRKHTGIYKIIAIVNEHQDEATVEVNILAPPSKPRGLDVKDVTKDS 1164
QY 98 ----- 97
DB 1165 CKLKKWKPEDDGGKPIISAYQVEKFKQGRWVPLGRTSANDTEFVKLGLEGHEYQFRVK 1224
QY 98 -----GPIQNY 103
DB 1225 AINEGESPLDSDSIIAKNPYDAASKPGTPNIVDYNEHMKLKWAEPRSDGAPISGY 1284
QY 104 SI-----TYLMFDFYSTQ----- 116
DB 1285 IIEKKDKPSIWDIELSTNTSVPEATVEGLVEGNIYQFRVAVNKAQSPSPDATEPHLA 1344
QY 117 ----- 116
DB 1345 KPRNLKPYINRDKMKPIKVRAGQPVDFVDVKGEPAPSLTWFLKTELTSTGTQVRLNID 1404
QY 117 -----LRKPAK----- 122
DB 1405 YNTKLTLLDTRDKSQSQYKLRANENINGVDEAVVEVILDKPSRKEGPIEVSDIHKEGCKL 1464
QY 123 -----YVYSQY-----NH----- 130
DB 1465 KWRKPKDDGGIPITGVVIEKMDTATGKWPAGSVDEPEKYDIEIKGLDPNHRIOFRVKAVN 1524
QY 131 -----TAKT-----ITFRPP----- 140
DB 1525 EGESEPLETESAITAKNPFDVSAPPGLPELEDWDEHHVKKLKWEPPIRDGSPITNYIE 1584
QY 141 ----- 140
DB 1585 VMDKSGEFKAVETDSPVCKGVVKKLEGGQQYKFRVAVNKAQSPSPQTNWHVAKPR 1644
QY 141 ----- 140
DB 1645 FLKPHIDRVNLKPIVKTGLSISLDINIRGEPAPKVEWFNNSSVTSDEHSVKIDNVYN 1704
QY 141 ----- 140
DB 1705 TKFFVMRAQRSOSGKYIIKATNEVGDEAELEVTVLGKPGKPLQVNDITKHSCKLKW 1764
QY 141 -----PCGR----- 144
DB 1765 EKPDGSGSPIDYIEIEKLDPHGTQWLPCGKSTPEBAKVIHLHGKAYKFRVAVNKEGE 1824

QY 145 ----- 144
Db 1825 SEDLETEKPIIAKNPYDEPDPRGKBPPTNWDKDFVDLAWPPKNDGGAPTKYVIQMRDK 1884
QY 145 -----VPSMTC----- 150
Db 1885 SGRAWDSATVPDGKCMGTGTGVBEGHEYPRIIVAVNKAAGSDPSDVSIAKPRELKP 1944
QY 151 -----LSEMLN----- 156
Db 1945 HIDRKNLOKIMRSGQMLHDALIKAEPPAKVVTYNTKTEIKSDHIKIENEDYKTTFIM 2004
QY 157 -----VSKRNDTG----- 164
Db 2005 PKVBRADRGIVITAKNDGSDTVEVEVLCKSPKPGPLAYSNVTAEHLHLKWKEPED 2064
QY 165 ----- 164
Db 2065 DGGDPIEQYLVERMDTETGRWPVLTTKTPREADVTGLTEGKEYLFRKAVNSEGESEPLV 2124
QY 165 ----- 164
Db 2125 TDIPTKAKNFDAAADTPGKQIIVDWSGNHCDLKWRAPEDDGGASITGYIIVERKDPNTGKW 2184
QY 165 ----- 164
Db 2185 QKALETSTPCKARVNDLIAKNKYQPRIAMAVNKAAGSKPSEPSDQMTAKDRFAPPKIDRT 2244
QY 165 ----- 164
Db 2245 NIKDITIKAGQHIRFDIKVSGEPPATKWVLHNLKARLENDSDNIDMESYRTKLTVPISK 2304
QY 165 -----EQCG----- 169
Db 2305 RFHSGKYTLKAENESGRDEASFEVIVLDKPPGPEPLRVTDVHKEGCKLKNWAPLDGGGL 2364
QY 170 ----- 169
Db 2365 PIDHYIIEKMDVESGRWLPSPGRKESFAELNNLEPSHEVKFRVLAVNTEGESEPLTGEQS 2424
QY 170 ----- 169
Db 2425 VIAKNPDEPGKGTPEAVDWDKHDVLDVWRPPINDGGSPITGYVVEKREKGTDKWIKGT 2484
QY 170 ----- 169
Db 2485 EITIPCLGEBCKATVPTLNENCEYEFVRKAINAAGPEPSDASKPIITKPKLAPKIDRK 2544
QY 170 -----NFTFNPMFF-----NVPWNTKLYV----- 190
Db 2545 NIRTYNKSGEPIFLDINISGEPAPOVTWNQNNKSVQTTSFSHIENLP-YNTK-YINNNP 2602
QY 191 -----GPTKV----- 195
Db 2603 ERKDTGLYKISAHNFYQDQVEFOIINIITKPGKPEGLEVEVHKGCKLKNWPKKDDGG 2662
QY 196 -----NVD-----SQTIFYLG 206
Db 2663 EPVESYLVEKFPDPTGTGLVPGRSDGPEYNVDGLVPGHDYKFRKAVNKEGESEPLETLG 2722
QY 207 ----- 206
Db 2723 SIIAKDPFSVTPKPGVPEPTDWTANKVELAWPEPASDGGSPIQGYIVEVKDKYSPLEWKA 2782
QY 207 ----- 206
Db 2783 LETNSPTTATVQGLIEGNEYQPRVALNKGGLSEPSDPSKIPTAKPRVLAPKIDRNLNR 2842
QY 207 -----LTALLRYAQRNCT 220
Db 2843 NITLSSGTALKLDANITGEPAPKVEWKNVHLSQGNVTIETPDYVTKVIRPTQRSDS 2902

QY 221 HSPYLVNA----- 228
Db 2903 GE-YLVTATNTSKDSSLVNVNVTDKPSPNGPLQISDVHKBGCHLKWKRPSDDGTPIE 2961
QY 229 -----MS 230
Db 2962 YFQIDKLEPETGCWIPSCRSTEPQVDVTGLSPGNEYKFRVSAVNAEGESQPLVGDESIVA 3021
QY 231 RNLFRYP-----KY----- 239
Db 3022 RNFDFEPGKPNLKAATWDKHDVLDLAWTPPLIDGGSPISCIYIIEKQDKYKWKERALDVPA 3081
QY 240 ----- 239
Db 3082 DQCKATIPDLVBEGQTYKFRVSAVNAAGTGEPSDSTPIIAKARNKPPIIDRSSLVEVRK 3141
QY 240 -----INGTKLKNMTMRKLKRKQ----- 256
Db 3142 AGQSFTFDCKVSGEPAPQTKWLLKKKEVYSKDNVKTNDVNTKLKVNSTRSDSGIYTV 3201
QY 257 ----- 256
Db 3202 FAENANGEDSADVKTVIDKPAPPNGPLKVDBINSECTLHMNPPDDGGQPIDNTYVVEK 3261
QY 257 -----APV-----KEQFEKKAKTOSTTTPYFYSYTTSAAL--- 286
Db 3262 LDETTGRWIPAGETDGPVTALKVGGLTPGHKYKFRVAKNRQGTSEP---LTTAQAIIAK 3318
QY 287 -----NVTT 290
Db 3319 NPFVDTPGTPTIKDFKFEVDLEWTRPEADGGSPITGYVVEKDKFSPDWEKCAEISD 3378
QY 291 NVT----- 293
Db 3379 DITNAHVPLDIEGLKYEFVRVAVNKAAGSPSDATETHVARPKNTPPKIDRNFMSDIKT 3438
QY 294 -----YSI 296
Db 3439 AGNVFEFDPVVTGEPPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILDAKESDGTGYTL 3498
QY 297 TTAARRVS----- 304
Db 3499 T--ARNINGTDRHNKVTILDAPSVEGPLRNGDVSKNSIVLWRPPKDDGGEITHYVV 3556
QY 305 ----- 304
Db 3557 EKQDNEMRWVPVGDCTDEIRADNLIENHDYSFRVAVNKGQSQPLTTSQPIITAKDPY 3616
QY 305 ----- 304
Db 3617 SHPDKPGOPQATDWGKHFDVLEWSTPKRDGGAPISSYIIEKRPKFGQWERAADVLDNCK 3676
QY 305 -----TSTIAYRP-----DSFPMKSI----- 320
Db 3677 AHVPELTNGEYEFVRVAVNRRGSPSDPSSTIICPRFLAPFPDKSILLANDITVHAGKR 3736
QY 321 -----MATQRLDLATWY----- 333
Db 3737 LGWTLPIEASPRPLITWLYNGKEIGSNRSGESGLFQNELTPEIVSSLSRSDGRTYLLIKN 3796
QY 334 ----- 333
Db 3797 EHGSPASAHATVLDLRPPKGPLDITKITRDGCHLTWNVPDDGGSPILHYIIEKMDS 3856
QY 334 -----TTLAYR-----QNPFCEP--- 346
Db 3857 RSTWSDAGNMSTHIVHDVTLVRKEYLFRKAVNAIGESDPLEAVNTIIAKNEFDEPDAP 3916
QY 347 -----SRNRVAVSEFM-----KNT----- 360
Db 3917 GKPIITDWRDHDLDQWAVPKSDGGAPISEYIIQKKEKGSYWTNVRHVPSKNKTTIPE 3976
QY 361 ----- 360

Db 3977 LTEQBVEFRVIAVNAQOSEPSESDMINAKPRYLPKIIITPLNEVRIRKCGLI FHTDIH 4036
QY 361 -----HVLIRN----- 366
|:|:|
Db 4037 FIGEPAEATWTLNSNPLLSNDRSTITSIGHSHVVHTVNCORSDSGIYHLLANS SGIDE 4096
QY 367 ----- 366
Db 4097 GSPFVVLDRPPEGPEMEYEETANSVTISWPKPKONGGSEISSVYIEKRDLTHGGGWV 4156
QY 367 ----- 366
Db 4157 PAVNYSAKYNHAVPRLEGTWYELRVMAENLQGRSDPLTSQPVVAKSQYTVPGAGK 4216
QY 367 ----- 366
Db 4217 PELTSDSKNHTIKWKOPISNGSGPIIGYDIERRDYNTGRWIKINGOPVPTAEYQDDRVT 4276
QY 367 ----- 366
Db 4277 SNHQYVRISAVNAAGNGKTSEPSAIFNARPLREKPRFYPDGLIGKRIKVRAGEPVNLNI 4336
QY 367 ----- 366
Db 4337 PISGAPTPTIEWKRGDLKLEBGKRISYETNSERTLFRIDDSNRDRSGKYTVTAANFEGKD 4396
QY 367 ----- 366
Db 4397 TADIEVIVVDKPSPPGEGPLSYTETAPDHISLHWSPKDDGSDITGYIIEFTBFGVDDWK 4456
QY 367 ----- 366
Db 4457 PVPGTCPNTNFTVKNLVEGKKYVFRIRAEINIYGASEALEGKPVAKSPDPGAPSOPTI 4516
QY 367 -----ETPYTI----- 372
Db 4517 SAYTPNSANLEWHPDDCGKPTIGYIVERRRERGGEWIKCNNTPTNTSYTNSLRD GAR 4576
QY 373 ----- 372
Db 4577 YBPRVLAVNEAGPHSKSPDPMTAEHQRYRPPDPBPDPDRITRNGVTLSWRPPTDGK 4636
QY 373 ----- 372
Db 4637 SRIKGYVEMRPNKGDKWKTNDIPINSTVYTPVSLKEGSEYSFRVVAENEGVGRSPSKP 4696
QY 373 ----- 372
Db 4697 SQPITIEQPNKPCMBELKVRDIVCRAGDDFSIHVPYLPAPKPNAFWYSNDNMLDDNNRV 4756
QY 373 ----- 372
Db 4757 HKHLTDDAASVVVNSKRADSGQYRLQKNTSGFDATINVRVLDPRSPPTRLRADEFSG 4816
QY 373 ----- 372
Db 4817 DSI LTYWNPNDGGSAIQNYIIEKKEARSSTWSKVSSCTVFPVVRIRNLVLNKEYDFRV 4876
QY 373 -----YG----- 374
|:|:|
Db 4877 IAKNKYGQSDPANTSEBILARHPDIPNTPGIPHGIDSTEDSITIAWTKPKHGGSPITG 4936
QY 375 ----- 374
Db 4937 YIIEKRLLSDDKWKVAHALCPDLCKIPNLINAEYFRVAAVNAAGSAYSGSSDLIF 4996
QY 375 -----TLDMS----- 379
|:|:|
Db 4997 CRPPHAPKITSLSIRDMTVIAGDEFIRITVPYHASPRTASWLSNGLLEVIPGERIKFDS 5056
QY 380 -----SLYYNET-----MFVENK-----TAS----- 395
|:|:|:|

Db 5057 NDYASMYNKSARKDETGSYTTITLTNNKSDTASCHVTVVDRPLPPOGPLNAYDITPDT 5116
QY 396 ----- 395
Db 5117 TLAWKTPLDGGSPITNYVVVEKLDNCSWMYKISSFVRNTHYDVNGLEPHYKYNFRVRAEN 5176
QY 396 -----DSNKT----- 401
|:|:|
Db 5177 QYGLSDPLDIIEPIVAKHQFTVPDEPGQPKVIDMDSGNVTLIWTRPLSDGSGRIQYQIE 5236
QY 402 -----PTSPSMGFQ----- 410
|:|:|:|
Db 5237 YRDILNDSSWNAVDYIIKDTKYQLYNLINGSYEYEFRIKAKNAAGLSKSPSPSLRFLKKG 5296
QY 411 -----RTFID----- 415
|:|:|
Db 5297 FTVPSPGAPQVTRVGKNYVDLKWKPLRDGGSRITGYIIERRDIGGAVWVKNDYVLD 5356
QY 416 ----- 415
Db 5357 TEYTVNMLJEMGDYEFPRFVAVNSAGRSBPSLCTMPIKVCVGLGGKPKDMITRLQDKVAPF 5416
QY 416 ----- 415
Db 5417 GKDYTLQCAASGKPSPTARWLNRNGKEIQMNGGRMTCDSDGVFRLHISNVQTDGDDGYTC 5476
QY 416 ----- 415
Db 5477 EAMNSLGFTVNTSGYLKIGSPPIINRCPSSELKLPEDGNSKIKIFYSGDQPLTVLKKNEV 5536
QY 416 ----- 415
Db 5537 ICDSNDDTHVKVNI FDDYVAIYIRNIVKSGGPGYQIEFTNESGSATGEFYVHITGMSAP 5596
QY 416 ----- 415
Db 5597 TQPMGISYINKNSCMLNWRPPSYDGLKVSHYVIERKDVSSPHWITVTSSTCKDTAFNVQ 5656
QY 416 ----- 415
Db 5657 LIENQEVIFRVMVAVNENGMPLEGLNPIRAKDPIDPPSPGSPQITEIGDGFVHLEWEK 5716
QY 416 ----- 415
Db 5717 PESDGAHIQGYWIDKREVGSNTRVNATICAANOINCINLIEGRQYEFRIFAQNVAGL 5776
QY 416 ----- 415
Db 5777 STESSASQAVKIIDPOAASPLIVKPLRDANCIONHNAQFTCTINGVPKPTISWYKGARE 5836
QY 416 ----- 415
Db 5837 ISNGARYHMYSEGDNHFLINDVFGEDADEYVCRAVNAKAGAKSTRATLAIMTAPKLNVP 5896
QY 416 ----- 415
Db 5897 RFRDTAYFDKGNVVIKIPFTGPKPRIHWRVDGENIESGGHYTVEXKHAHLIIRDGS 5956
QY 416 -----PLWD----- 419
|:|:|
Db 5957 HLDGSPYRITAENELGSDTAIIQVQISDRDPDRPPPLIESIGTESISLSWKAPVMDCCSD 6016
QY 420 -----YL----- 421
|:|:|
Db 6017 ITNYVVEREHPPLSSWIRVGNTRFTSMVAVSGLTPGKEYDFRIPADNVYGRSDASDTSLI 6076
QY 422 -----DSLIF----- 426
|:|:|
Db 6077 KTKESVKKKPIERKWEIDANGRKLRGADGPKVDYDYVFDIYSKFVQPVVEISQSVYD 6136
QY 427 -----LDEI----- 430
|:|:|
Db 6137 RYDILEEICTGAPGVVHRCRERSTGNI FPAKFIPVSHSVKDLIRREIDIMNQLHOKLI. 6196

QY 431 ----- 430
Db 6197 NLHDAPEDDDEMILILEFLSGGELFERITAEGYVMTAEVINYMQICEGIRHHEQNII 6256
QY 431 ----- 430
Db 6257 HLDIKPENIMCQTRSRSTNVKIIDFGLATRLDPNEVKITTTGTAFAFAPEIVNREPVGFT 6316
QY 431 ----- 430
Db 6317 DMWATGVLSYVLLSGLSPFAGDNDVQTLKNVACDMDVDFESFKYISEEAKDFIRKLLVR 6376
QY 431 ----- 430
Db 6377 NKEKMTAHECLLHPWLTDHSAMKQEIINRDRVIAEYREKLRRKYEDPERFLLPIGRLSEY 6436
QY 431 ----- 439
Db 6437 SSLRKLIMEKXKIHDAVDFRRQAAPREVIRPSSQFCYEQSVKFCYRCIAIATPTLTWSH 6496
QY 440 ----- 445
Db 6497 NNIELQSVKFMKRYVGDDYFIINRVKLDORGEYIIRAENHYGSRREVVFLNVQPLPK 6556
QY 446 ----- 450
Db 6557 QPRYRTESTFVRREPLPYTFWQEESETPSPFTLLRPRVMQARDCKLLCCLSGKPVN 6616
QY 451 ----- 450
Db 6617 VRWYKDGRELSKEYAMTHSDGVVTWEIIDCKPDSGSKYCATNCHGTDETCVVIVEG 6676
QY 451 ----- 456
Db 6677 EWTPEQAQLAHNPLYSGDRKYIEQPIKPAPLPIVTSROYTSSVQNTSEPOGDKVNSN 6736
QY 457 SNS ----- 459
Db 6737 SNSSGISNKKKYASNSIQAPGSPSRSRATKELLIPDDSLMCKPEFTKPLHDLTIHDGE 6796
QY 460 ----- 459
Db 6797 QLILTCYVKGDBPQISWSKNGKSLSSDILDLRYKNGIATLTINEVFPDEGVITCTAT 6856
QY 460 ----- 459
Db 6857 NSVGAVETKCKLTIQPLDKNINKRVNAGDNAPKIVSHLESFRVRDGDVNLACRIIGAQ 6916
QY 460 ----- 462
Db 6917 HPDVVLHNNKEIKPSKDPQYTNEMNIYRLQIAEIPFEDGGTYTCEAFNDIGESFSTCTI 6976
QY 463 ----- 465
Db 6977 NVTVPGETKQPSFVKPFTSVSVLEGGTTFECEIDSELINLVWLK 7022

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:33:27 ; Search time 40 Seconds
(without alignments)
1118.520 Million cell updates/sec

Title: US-09-942-146A-1

Perfect score: 2451

Sequence: 1 MGRKEMVVRDVPKMFVLISI.....PEHRAVNLSNSLWMLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2417	98.6	466	2 S09837	hypothetical prote
2	681	27.8	26926	1 I38344	titin, cardiac mus
3	575	23.5	15281	2 S41309	cyclosporin synth
4	555	22.6	10797	2 T30192	probable peptide s
5	520	21.2	8243	2 T31307	type I fatty acid
6	500	20.4	6658	2 T13931	projectin - fruit
7	500	20.4	9376	2 T14593	syngomycin synth
8	498	20.3	7829	2 T15789	hypothetical prote
9	479	19.5	13055	2 T16580	hypothetical prote
10	473	19.3	7160	2 T27935	hypothetical prote
11	472	19.3	5105	2 T32650	hypothetical prote
12	471	19.2	7962	2 I38346	elastin titin - hu
13	469	19.1	6805	2 S20901	titin - rabbit (fr
14	469	19.1	6831	2 A88852	protein unc-22 (im
15	469	19.1	6839	2 S27242	twitchin [similar
16	465	19.0	4688	2 F82885	hypothetical prote
17	465	19.0	5005	2 F82884	hypothetical prote
18	465	19.0	10223	2 T30225	polyketide synthas
19	458	18.7	6669	2 S55024	nebulin, skeletal
20	452	18.4	5825	2 T12117	polyprotein - fava
21	451	18.4	4924	2 T50176	probable peptide s
22	450	18.4	6359	2 T31679	bacitracin synthet
23	449	18.3	4563	1 LPHUB	apolipoprotein B-1
24	449	18.3	6486	2 T31076	tyrocidine synthet
25	447	18.2	8563	2 T30226	polyketide synthas
26	444	18.1	5147	1 IJEFPM	cadherin-related t
27	443	18.1	4466	1 S17231	dynein beta heavy
28	443	18.1	4466	1 S17653	dynein beta heavy
29	443	18.1	4589	2 T14914	dynein beta heavy

30	443	18.1	4767	2 T31345	hypothetical prote
31	441	18.0	5232	2 A45086	HC-toxin synthetas
32	441	18.0	5369	2 T44807	mycosubtilin synth
33	440	18.0	4644	1 A38905	dynein heavy chain
34	437	17.8	4436	2 E71086	hypothetical prote
35	437	17.8	7576	2 T17428	FK506 polyketide s
36	436	17.8	5255	2 T31677	bacitracin synthet
37	435	17.7	3791	1 YGFLV8	alpha-aminoadipyl-
38	435	17.7	4588	2 T28667	dynein beta heavy
39	435	17.7	4930	2 E69679	polyketide synthet
40	434	17.7	4092	1 S38128	dynein heavy chain
41	434	17.7	4196	2 T43274	dynein heavy chain
42	433	17.7	5138	2 E96695	hypothetical prote
43	432	17.6	4485	2 T08044	dynein gamma heavy
44	431	17.6	4725	1 A44357	dynein heavy chain
45	431	17.6	4910	2 S64942	probable membrane

ALIGNMENTS

RESULT 1

S09837

hypothetical protein UL74 precursor - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5

A:Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004

C:Accession: S09837

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Hornsneil, T.;

M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; PMID:90269039; PMID:2161319

A:Accession: S09837

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-466 <CHE>

A:Cross-references: UNIPROT:P16750; EMBL:X17403; NID:G59591; PIDN:CAA35389.1; PID:G17808

A:Note: this sequence was submitted to the EMBL Data Library, December 1989

C:Superfamily: human cytomegalovirus hypothetical protein UL74

C:Keywords: glycoprotein

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-466/Product: hypothetical protein UL74 #status predicted <MAT>

F:75,83,87,103,130,157,162,171,219,242,288,292,350,367,385,392,399,433,443,454/Binding s

Query Match 98.6%; Score 2417; DB 2; Length 466;

Best Local Similarity 98.9%; Pred. No. 3.9e-60;

Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MGRKEMVVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60

Db 1 MGRKEMVVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRLGLVLSKIGYKLDQLKL 60

Qy 61 EILRQLETTISTKYNVSKQPVKNLTWN-TRPPQYIILAGPIQNYISITLWDFSTQLRK 119

Db 61 EILRQLETTISTKYNVSKQPVKNLTWN-TRPPQYIILAGPIQNYISITLWDFSTQLRK 120

Qy 120 PAKVYVSYQYVHTAKTITFRPPQGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179

Db 120 PAKVYVSYQYVHTAKTITFRPPQGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180

Qy 180 NVPRWNTKLVGPTKKNVDSQTIYFLGLTALLLYAQNRCTHSPFLVNSRNLFVRPKY 239

Db 180 NVPRWNTKLVGPTKKNVDSQTIYFLGLTALLLYAQNRCTHSPFLVNSRNLFVRPKY 240

Qy 240 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPTTFSTTSAALNVTNTVYSITTA 299

Db 240 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPTTFSTTSAALNVTNTVYSITTA 299

Qy 241 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPTTFSTTSAALNVTNTVYSITTA 300

Db 241 INGTKLNTWRKLRKQAPVKEQEKAKKTQSTTTTPTTFSTTSAALNVTNTVYSITTA 300

Qy 300 ARRVSTSTIAYRPDSSFMKSMATQLRDLATWVTTTLRYRQNPFCEPSRNRTAVSEFMKN 359

Db 300 ARRVSTSTIAYRPDSSFMKSMATQLRDLATWVTTTLRYRQNPFCEPSRNRTAVSEFMKN 360

Qy 360 THVLRNETPTVITYGTLDMSSLYNETMFVBNKTASDNKTTPTSPGMSGQRTFIDPLWD 419
Db 361 THVLRNETPTVITYGTLDMSSLYNETMFVBNKTASDNKTTPTSPGMSGQRTFIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSLSRPTVYVNLTPPEHRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSLSRPTVYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 2
I38344
titin, cardiac muscle [validated] - human
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R:Labeit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Accession: I38344
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Triatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:9602580
A:Note: conformation and properties are reported for a synthetic peptide corresponding t
R:Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'P', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2
A:Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R:Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A:Title: Genomic organization of M line titin and its tissue-specific expression in two
A:Reference number: S63665; MUID:96177761; PMID:8604138
A:Accession: S63665
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 26729-26825 <KOL>
A:Cross-references: EMBL:X92412; NID:g1236761
R:Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A:Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia
A:Reference number: S37393; MUID:94008990; PMID:8404852
A:Accession: S37393
A:Molecule type: mRNA
A:Residues: 26831-26926 <GAU>
R:Improta, S.; Pollton, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996

A:Reference number: A66736; PDB:1TIT
A:Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R:Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A:Reference number: A66201; PDB:1NCT
A:Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C:Genetics:
A:Gene: GDB:TTN
A:Cross-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q32
C:Function:
A:Description: structural protein forming filaments in striated muscle
C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C:Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F:24752-25008/Domain: protein kinase homology <KIN>
F:84,177,905,2276,2378,2459,2481,2563,2569,2763,2896,3088,3179,3384,3432,3628,3772,4068,
98,11066,11486,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F:16780,16976,17579,17602,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
.21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F:26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental
Query Match 27.8%; Score 681; DB 1; Length 26926;
Best Local Similarity 1.0%; Pred. No. 0.00038;
Matches 268; Conservative 91; Mismatches 106; Indels 25559; Gaps 84;
Qy 1 MGRK----- 4
Db 649 VGKAEAVATVAANDQAVRREPFGHLESVAQQTLLGYGYKERISAAKVAEPQPPA 708
Qy 5 ----- 4
Db 709 SEPHVVPKVPKRVIQAPSETHIKTTDQKGMHSSQIKKTTDLTLRLVHDKRPRTASP 768
Qy 5 ----- 4
Db 769 HFTVSKISVPKTEHGYEASIAIATFLQKLSATSAQKITKSVKAPTVPKSETRVRAE 828
Qy 5 ----- 4
Db 829 PTPLPQFPFADPTPYKSEAGVEKVGVSITCTTVREERFVHLHGREAKVTETARVPA 888
Qy 5 ----- 4
Db 889 PVEIPVTPPTLVSLGNVTVEGESVTLECHISGYPSTVTWYREDYQIESSIDFQITPQ 948
Qy 5 ----EMVRD----- 10
Db 949 SGIARLMIREAFADSGRFTCSAVNEAGTSTCYLAVQVSEEFKEKETTAVTEKFTTEEK 1008
Qy 11 ----- 10
Db 1009 RFVESRDVMTDTSLTEEQAGPGEPAAPIFTKPVQKLVGGSVVFGCGVGNPKPHVY 1068
Qy 11 ----- 10
Db 1069 WKSGVPLTTGYRYKVSYNKQTGECKLVISMTFADDAGEYTIIVVRNKHGETSASASLEE 1128
Qy 11 ----- 10
Db 1129 ADYELLMKSOEMLYQVTFATVQEPVEGTAPGFTVSEYEKEYEKOALIRKMAKDTV 1188
Qy 11 ----- 10
Db 1189 VRTVVEDQEPHISFEERLIKIEYRIKTTLELLEDGEKMAVDISEAVESEGFD 1248
Qy 11 -----VPK----- 13
Db 1249 LRKNYRILEGMVTFHCMSGYPLPKIATWYDKGRINHGERYQMDFLQDGRASLRIPVV 1308
Qy 14 ----- 13

```
Db 1309 LPEDEGIYAFASNIKGNACSGKLYVEPAAPLGAPTYIPTLEPVSRIRSRSPSRSP 1368
Qy 14 -----MEVLISISFLVSPFNC 30
Db 1369 IRMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMSPARMS 1428
Qy 31 -----KWSK-----35
Db 1429 RFDLKVGRRPMPETFWFHDGOQIVNDYTHKVVIKEDGTQSLIIVPATPSDGEWTVVAQN 1488
Qy 36 -----35
Db 1489 RAGRSSISVILTVEAVEHQVPMFVEKLVNVIKEGSRLEMKVRATGNPNPDIVWLKNSD 1548
Qy 36 -----35
Db 1549 IIVPHKPKIRIEGTGKAALKIDSTVSDSAWYTATAINKAGRDTTRCKVNVVEPAEP 1608
Qy 36 -----35
Db 1609 EPERKLIIPRGTYRAKEIAPELEPLHLRYGOEQWEGDLYDKSKQKPPFKKLTSLRL 1668
Qy 36 -----35
Db 1669 KRFGPAHFECRLTPISDPTMVVWVWLHDGKPLEAANRLMINEFGYCSLDYGVAYSRDSGI 1728
Qy 36 -----35
Db 1729 ITCRATNKYGTHTSATLIVKDEKSLVESQLPEGRKGLQRIEELERMAHEGALTGVTTD 1788
Qy 36 -----ALYNRPWR-----43
Db 1789 QKEQKQPDIVLYPEVRVLEGETARFCRVGTGYPOPKNVWYNGOLIRKSRFRVYDGI 1848
Qy 44 -----GLVLSKI-----50
Db 1849 HYLDIVDCKSYDTEGVKVTAEENPEGVIEHKVKLEIQOQREDPRSVLRRAPEPRPFHVHEP 1908
Qy 51 -----50
Db 1909 GKLPQEVQKVRPVDTTTETKEVVKLKRAERITHEKVPSEBELRSKFKRTEEGYEAIT 1968
Qy 51 -----50
Db 1969 AVELSKRKDBSYELLKTKDELHMTKELTEBEKKALABEGKITPTFKPKDKIELSPS 2028
Qy 51 -----50
Db 2029 MEAPKIFERIOSQTVGQSDAHFRVVRVVGKPDPECEWYKNGVKJERSDRIIYWPEDNVC 2088
Qy 51 -----50
Db 2089 ELVIRDVTAEDSASIMWKAINAGTSSHAFLLVQAKOLIITFTQELQDVVAKEKDTMATP 2148
Qy 51 -----50
Db 2149 ECETSEPFVKWKYKMGVEHGDYRMHSDRKVHFLSILITDTSADYSCVLVEDENV 2208
Qy 51 -----GKYKL 55
Db 2209 KTTAKLIVEGAVFVKELQDIEVPESYSGELCIVSPENIEGKWHYHNDVELKNGKTYI 2268
Qy 56 -----DOLK 59
Db 2269 TSRRGRQNLTKVDVTKEQGEYSFVIDGKTKTKLKMKPRPIALLOGLSDOKVEGDIVQ 2328
Qy 60 LEILRQLET-----TI-----70
Db 2329 LEVKVLSSEVGWMKQGEVQPSDRVHVIDKQSHMLLIEDMTKEDAGNYSFTIPALGL 2388
Qy 71 -----70
Db 2389 STSGRVSVSYVDVITPLKDVNVNIEGTKAVLECKVSPDVTSVKRWYLNDEQIKPDDRVOAI 2448
```

```
Qy 71 -----70
Db 2449 VKGTQORLVINRTHASDEGPYKLIIVRVTNCNLSVEKIKIIRGLRDLTCTETQWVPEV 2508
Qy 71 -----STKYNVS-----77
Db 2509 ELSHSGIDVLNFKDKKEIKPSSSKYKIEAHGKIYKLVNLMMKDDGKTYTFVAGENWTSKG 2568
Qy 78 -----77
Db 2569 LTVAGGALSPLDTQTVABSOEAVPECEVANPDGKGLPLTLNNIRSESDGHR 2628
Qy 78 -----KQPVKNLTM-----86
Db 2629 RLIIAATKLDIDIGYTYKVATSKTSAKLVBAVKIKTKLKNLTVTETQDAVFTVELTHPN 2688
Qy 87 -----86
Db 2689 VKGVOMIKNGVWLESNEKYAISVKGTIYSLRIKCAIVDSEVYFRLGRLGASARLHVET 2748
Qy 87 -----86
Db 2749 VKIIKPKDVTALENATVAPEVSYSHTVPVKWPHKSVEIKPSDKGRLVSRKVKHKLMLQ 2808
Qy 87 -----NTEFPQ-----92
Db 2809 NISPSDAGEYTAVVQGLECKAKLFVETLHTTKTKMKNIEVPETKTASFECEVSHFNVPSSM 2868
Qy 93 -----92
Db 2869 LKNGVEIEMSEKPIVVQGLHOLIIMNTSTEDSABYTFVCGNDQVSATLTVTPIMTSM 2928
Qy 93 -----92
Db 2929 LKDINASEKDTITFEVTVNYEGISYKWLKNGVEIKSTDKCOMRTKLTSLNIRNVHFGD 2988
Qy 93 -----92
Db 2989 AADYTFVAGKATSTATLYVEARHIEFRKHIDIKVLEKKRAMFECEVSEPDITVQMMKDD 3048
Qy 93 -----YILAG-----98
Db 3049 QELQITDRIKIQEKYVHRLIIPSTRMSDAGKYTVVAGGNVSTAKLFVEGRDVRIRSICK 3108
Qy 99 -----98
Db 3109 EVQVIEKQRAVVEPEVNEDDVDAHMYKDGIENFQVQERHKYVVERRIHRMFISETROSD 3168
Qy 99 -----98
Db 3169 AGEYTFVAGNRSSVTLYVNAPEPPQVLQELQPVTVQSGKPARFCAMISGRPQPKISWYK 3228
Qy 99 -----98
Db 3229 EEQLLSTGFKCKFLHDQOEYTLILLIEAFPDAAVYTCEAKNDYGVATTSASLSVEVPEVV 3288
Qy 99 -----PLQN-----102
Db 3289 SPDQEMPVYPPIITPLQDVTVTSEGQPARFCRVSGTDLKVSWYSKOKKIKPSRPFRTMQ 3348
Qy 103 -----102
Db 3349 PEDTYQLEIAEAYPEDESGTYTFVANNVAGQVSSSTANLSLEAPESILHERISQEIEMEMKE 3408
Qy 103 -----102
Db 3409 FSSSFLSABEGLHSLQLSKINETLELLESVPYPTKFDSEKGTGPIFIKEVSNADI 3468
Qy 103 -----102
Db 3469 SMGDVATLSVTVIGIPKPKIQWFFNGVLLTSPSADYKVFVFDGDDHSLIILFTKLEDEGEYT 3528
```

Qy 103 ----- 102
 Db 3529 CMASNDYKTTICSAVLKINSRGEHGKHDTETESAVAKSLKLGPCPPHFLKELPIRCAQ 3588
 Qy 103 -----TVLWF----- 110
 Db 3589 GLPAIFETVVGEPAPTWTFKENKQLCTSVYTTIIHNPNGSGTFIVNDPOREDSGLYIC 3648
 Qy 111 -----DFVSTOLRKA----- 121
 Db 3649 KAENMLGESTCAAEALLVLEDDTMDTPCKAKSTPEAPEDFPQPLKGPFAVEALDSQEI 3708
 Qy 122 ----- 121
 Db 3709 ATFFVKDTILKAALITEENQQLSYEHAKANELSSQLPLGAQELQILLEQDKLTPESTREF 3768
 Qy 122 ----- 121
 Db 3769 LCINGSIHFOPLKEPSPNLQIIVOSQKTFSEKILMPEEBETOAVLSDTTEKIFPSAMSI 3828
 Qy 122 ----- 121
 Db 3829 EOINSLTVEPLKTLAEPEGNYPOSSIIEPPMHSYLTSVAEVLSEKKTVSdTNRQRVT 3888
 Qy 122 ----- 121
 Db 3889 LQKQEQSALILSQSLABGHVESLQSPDVMISQVNYEPLVPSEHCTEGGKILIESANPL 3948
 Qy 122 ----- 121
 Db 3949 ENAGQSAVRIEEGKSLRFPPLALBEKVLLKEEHSNDVVMPPDQIIESKREBFAIKVQE 4008
 Qy 122 ----- 121
 Db 4009 VQGRDLLSKESLLSGIPEEQRLNLKIQICRALQAQAVASEQGLPSEWLNRNIEKVEEAVN 4068
 Qy 122 ----- 121
 Db 4069 ITQEPRHIMCVLYTSAKSVTEEVTTIIEDVDVDPQMANLKMELDALCAIIVEEIDILTAE 4128
 Qy 122 ----- 121
 Db 4129 GPRIOQAKTSLQEBMDSFGSQKVEPITEPEVESKYLISTEESYFNVQSRVKYLDATP 4188
 Qy 122 ----- 121
 Db 4189 VTKGVASAVVSEKQDESLLKPSSEKESSESSGTEEVATVKIQEABGGLIKEGDMIHTP 4248
 Qy 122 ----- 121
 Db 4249 LVDIVSEBGDIVHLTTSITNAKEVNWYFENKLVPSDEKFCIQDQNTVTLVIDKVNTEDH 4308
 Qy 122 --KYVSYNNHTAKTI----- 135
 Db 4309 QGEYVCEALNDSGKTATSAKLTWVKRAAPVIRKIEPLEVALGHAKFTCEIQAPNVRF 4368
 Qy 136 ----- 135
 Db 4369 QWFKAGRIEYSDKCSIRSKYISLSLEILRTQVVDVCGEYTCASNEYSGVSCTATLFTVT 4428
 Qy 136 ----- 135
 Db 4429 PGGEKVKRKLPERKPEKEEVLKSVLRKPEBEEPKVEPKLEKVKKPAVPEPPPKP 4488
 Qy 136 -----TFRPPP----- 141
 Db 4489 VEEVEVPTVTKRERKIPEPTKVPEIKPAIPLPAPEPKPEAEVKTIKPPVPEPEPTPIA 4548
 Qy 142 ----- 141
 Db 4549 APVTVPVVGKAEAKAPKEEAAKPGPIKGVPKTTPSIEARRKLRPGSGGKPPDEAP 4608
 Qy 142 ----- 141

Db 4609 FTYQLKAVPLKFVKKEIKDIILTESEFVGSSAIFCELVSPSTAITWTMKGDSNIRESPKHR 4668
 Qy 142 ----- 141
 Db 4669 FIADGDKRKLHIIDVQLSDAGEYTCVLRGNKKEKTSTAKLVEELPVRFKVTLSEEVTVV 4728
 Qy 142 -----C-----GRV----- 145
 Db 4729 KGQPLYLSCELNKERVVWRKDGKIIVVEKPGRIVPGVIGLMRALTIINDADDDAGTYTVT 4788
 Qy 146 ----- 145
 Db 4789 VENANNECSCVKVVEVIRDWLKPIRDQHVKPGKTAIFACDIAKDTPNIKWFKGYDEI 4848
 Qy 146 ----- 145
 Db 4849 PAEPNDKTEILRDGNHLYLKIKNAMPEDIAAYAVEIEGKRYPAKLTIGEREVELLKPIED 4908
 Qy 146 -----PSMTC----- 150
 Db 4909 VTIYKESASFDAIESEADIPGQWKLGELLRPSPTCEIKABGGKRFLTLLHKVKLDQAGE 4968
 Qy 151 ----- 150
 Db 4969 VLYQALNAITTAILTIVKEIELDFAVPLKDVTVPERQARPECVLTREANVWSKGPDIK 5028
 Qy 151 ----- 150
 Db 5029 SSDKFDIIADGKXHLVINDSQFDDGVTYAEVEGKTSARLFTVGTIRLKPMSPLEDQTV 5088
 Qy 151 ----- 150
 Db 5089 KEGETATFVCELSHEKMHVWFKNDAKLHTSRTVLISSEGKTHKLEKVEVTLDDISQIKA 5148
 Qy 151 ----- 150
 Db 5149 QVKELSTAQLKLEADPYFTVKLHDKTAVEKDEITLKCEVSKDVPVKWFKDGEIIVPSP 5208
 Qy 151 ----- 150
 Db 5209 KYSIKADGLRRIILKIKKADLKDKGEYVDCDGTKTANVTVEARLIEVEKPLYGVEVPVG 5268
 Qy 151 ----- 150
 Db 5269 ETAHFBIELSEPDVHGQWKLKGQPLTASPDCEIIEDGKXHLILHNCQLGWTGEVSFQAA 5328
 Qy 151 ----- 150
 Db 5329 NAKSAANLKVKEPLFIITPLSDVKVFEKDEAKPECEVSREPKTFRWLKGTQBITGDORF 5388
 Qy 151 ----- 150
 Db 5389 ELIKDGTGKSNVTKSAFAFEDEAKYMFABDKHTSGKLIIEGIRLKLFTPLKDVTAKEKES 5448
 Qy 151 ----- 150
 Db 5449 AVFTVELSHDNIRVKWFKNDORLHTTRSVSQDEGKTHSITFFKDLISIDDTSQIRVEANGM 5508
 Qy 151 ----- 150
 Db 5509 SSEAKLTVLEGDPYFTGKLQDYTGVEKDEVILQCEISKADAPVKWFKDGKEIKPSKNAV 5568
 Qy 151 ----- 150
 Db 5569 KTDGKKRMLILKKALKSDIQOYTCDGTDKTSKGLDIEDREIKLVRPLHVSVEVMEETAR 5628
 Qy 151 ----- 150
 Db 5629 PETEISEDDIHANWKLKGEALLQTPDCEIKEEGKIHSLVHLNCRCLDQGTGGVDFOQANVKS 5688
 Qy 151 ----- 150

Db 5689 SAHLRVKPRVIGLLRPLKDVTVTAGETATDFCELSYEDIPVENVYLGKGLBPSDKVPRS 5748
QY 151 ----- 150
Db 5749 EGKVHTLRLDVKLEBAGEVOLTAQDKFKTHANLPVKEPPVEFTKPLEDQTVVEGATAVLE 5808
QY 151 ----- 150
Db 5809 CEVSRENAKVKWFKNGTILKSKYIEIVADGRVKLVIHDCPTEDIKTYTCDAKDFKTSK 5868
QY 151 ----- 150
Db 5869 NLNVPPHVEFLRLPLDLQVREKENARFECELSRENAKVKWFKDGAIEIKGKKYDIISKG 5928
QY 151 ----- 150
Db 5929 AVRILVINKLLDDEABYSCVVRTARTSGMLTVLEBEAVFTKNLANIEVSETDTIKLVCE 5988
QY 151 ----- 150
Db 5989 VSKGAEVIWYKDBEIIETGRYELTEGRKRILVIONAHLEDAAGNYNCRLPSSRTDGKV 6048
QY 151 ----- 150
Db 6049 KVHELAAEFISKPNLEILEGEKABFVCSISKESPPVQWKDDKTLESGDKYDVADGKK 6108
QY 151 ----- 150
Db 6109 RVLVVKDATLQDMGTVMVVGAAAHILTIEKLRIVVPLKDTFRVKEQEVVFNCEVNT 6168
QY 151 ----- 150
Db 6169 EGAKAKFRNBEAIPDSSKYIILQKDLVYTLIRDAHLDDQANVNSLTHRGENVKSAA 6228
QY 151 ----- 150
Db 6229 NLIVEEDLRIVEPLKDIETMEKKSVTFWCKVNRNLNVLTKWTKNGEEVPPFNRSYRVDK 6288
QY 151 ----- 150
Db 6289 YKHLMTIKDCGPDGSEYIVTAGQDKSVAELLIIIEAPTFFVEHLEDQTVTFDDAVFSCQ 6348
QY 151 ----- 150
Db 6349 LSREKANVWYRNGREIKEGKKYKPEKDGSIHRLIIKDCRLDDECEYACGVEDRKSARL 6408
QY 151 ----- 154
Db 6409 FVEEIPVEIIRPPQDILEAPGADVFLAELNKKVQVQWLNRRNMVVGDKHKQMSSEGI 6468
QY 155 ----- 154
Db 6469 HRLQICDIKPRDQGEYRFIANDKEARAKLELAAPKIKTADQDLVVDVGKPLTWVVPYDA 6528
QY 155 ----- 154
Db 6529 YPKAEAWFKENEPLSTKTIIDTBAQTSFRILEAKKGDKGRYKVLQNKHGKABGFNLK 6588
QY 155 ----- 154
Db 6589 VIDVGPVRNLEVTETFDGEVSLAWEBPLTDGGSKIIGYVVERRDIKRKTWVLATDRAES 6648
QY 155 ----- 157
Db 6649 CEFTVTGLQGGVEYLFVRSARNVGTGEPVETDNVPEARSKYDVPGLPLNVTITDVRNP 6708
QY 158 ----- 157
Db 6709 GVSLTWBPPPYGGAEITNYVIELRDKTSIRWDTAMTVRAEDLSATVTVVVEGQEYSFRV 6768
QY 158 ----- 157
Db 6769 RAQNRIGVKFSAATPFVKVADPIERPSPVNLTSDDQTSVQLKWEPLPLKDGSPILG 6828

QY 158 ----- 157
Db 6829 YIIBCEBGKDNWIRCNMKNLVPCLTYKVTGLEKGNKYLRYSAENKAGVSDPSEILGLPLT 6888
QY 158 ----- 157
Db 6889 ADDAPVEPTMDLSAFKOGLEVI VNPITILVPSTGYPRPTATWCFGDKVLETGDRVKWMT 6948
QY 158 ----- 157
Db 6949 LSAYAEVLVSPERSDKGIYTLKLENRVKTTISGEIDVNVNVIARPSAPKELKFCGDIITKDSVH 7008
QY 158 ----- 157
Db 7009 LTWEPDDGGSPLTGYVVEKREVSRTWTWKVMDFTDLBFTVDPDLVQGKEYLKFVCARN 7068
QY 158 ----- 157
Db 7069 KCGEPAYVDEPVNMSTPATVPDPENNVKWRDRTANSIFLTWDPKNDGGSRIKGYIVB 7128
QY 158 ----- 157
Db 7129 RCPRGSKWVACGEPVAETKMEVTGLBEGKYYAYRVKTLNRQGASKPSRPTTEEIOAVDTQ 7188
QY 158 ----- 157
Db 7189 EAPBIFLDVLLAGLTVKAGTKIELPATVTKGPEPKITWTKADMILKQDKRITIEVNPVK 7248
QY 158 ----- 164
Db 7249 STVTIVDSKRSDDTGYTII EAVNVCGRATAVVEVNVLDKPGPPAAPDITDVTNESCLLTWN 7308
QY 165 ----- 164
Db 7309 PPRDDGSKITNYVVERRATDSEVWHKLSSTVKTDFNPKATKLI PNKEYIFRVAENMYGA 7368
QY 165 ----- 164
Db 7369 GEPVQASPI TAKYQDPGPPPTLRLEPSDI TKDAVTLTWCEPDGSGSPITGVVVERLDPD 7428
QY 165 ----- 164
Db 7429 TDKVRCNKMPKDTTYRVKGLTNKKYRFRVLAENLAGPGKPSKSTEBILIKOPIDPPW 7488
QY 165 ----- 164
Db 7489 PPGKPTVKDVGKTSVRLNWTYKPEHDGGAIESYVVIEMLKGTGTDEWVRVAEGVPTQHLLP 7548
QY 165 ----- 164
Db 7549 GLMEGQEYSFRVAVNKA GSESPSPDVL CREKLYPPSPRWLEVINITKNTADLKWT 7608
QY 165 ----- 164
Db 7609 VPEKDGSPITNYIYEKRDVRRKGWQTVDTTVTKTCTVTPLTEGSLYVFRVAENAIGQ 7668
QY 165 ----- 164
Db 7669 SDYTEI EDSVLAKDTFTTTPGPYALAVVDVTKRHVDLKWBPKNKGGRPIQRYVIEKKER 7728
QY 165 ----- 169
Db 7729 LGTRWVKAGTAGPDCNFRVTDVIEGTEVQFQVRAENEAGVGHFSEPTTEILSIEDPTSP 7788
QY 170 ----- 169
Db 7789 SPPLDLHVTDAGRKHIAIAWKPEKNGSGPIIGYHVEMCPVGTCKWVRNRPINKDLKFK 7848
QY 170 ----- 169
Db 7849 VEGVVPDKKEYLVRVAVNAIGVSEPSSEISNVVAKDPDCKPTIDLETHDIIVIEGEKLS 7908

QY 170 ----- 169
Db 7909 IPVPFRAVPVTVSMHKGKEVKASDRLTMKNDHISAHLEVPKSVRADAGIYITILENKL 7968
QY 170 ----- 169
Db 7969 GSATASINVKVIGLPCKDIKASDITKSSCKLTWEPPEFDGGTPIHLHYLERREAGRR 8028
QY 170 ----- 169
Db 8029 YIPVMGENKLSWTVKDLIPNGEYFPRVAVKVNKGGEYIELKNPVIADPKQPPDPVD 8088
QY 170 ----- 169
Db 8089 VEVHNPATAEMTITWKPPLYDGSKIMGVITIEKIAKGEHMKRCNEHLVPILTYTAKGLE 8148
QY 170 ----- 169
Db 8149 EGKEYQFRVRAENAAGISEPSRATPPTKAVDPIDAPKVILRTSLEVKGDEBIALDASISG 8208
QY 170 ----- 169
Db 8209 SPYPTITWKENVIVPEIIEKRAAPLVRRRKGEVQEEEPVLPTQLRSIDNSKKGESQ 8268
QY 170 ----- 169
Db 8269 LRVRDSLRPDHGLYMIKVENDHGIAKAPCTVSVLDTGPPINFVPEDIRKTSVLCKWEPP 8328
QY 170 ----- 172
Db 8329 LDDGSEIINTLEKKDKTKPDSEWIVVTSLRHCKYSVTKLIEGKEYLFRVRAENRFGP 8388
QY 173 -----TFNPMF----- 178
Db 8389 GPPCVSKPLVAKDPGPPDADKPIVEDVTSNMLVKWNEPKDNGSPILGWLEKREVNS 8448
QY 179 ----- 178
Db 8449 THWSRVNKSLLNALKANVDGLLEGLTYFRVCAENAGPGKFSPPSDPKTAHDPISPPGP 8508
QY 179 ----- 178
Db 8509 PIPRVDTSTTIELEWEPAPNGGGEIVGYVDKQLVGTNKSRCRTEKMIKVRQYTVKE 8568
QY 179 ----- 178
Db 8569 IREGADYKLRVSANNAAGEPPGETQPTVAEPQEPAPVELDVSVKGIQIMAGKTLRIP 8628
QY 179 ----- 178
Db 8629 AVVTGRPVPTKVWTKGEGELDKDRVVIDNVGTKSELIIKDALKDGHGRYVITATNSCGSK 8688
QY 179 ----- 178
Db 8689 FAAARVEFDVPGPVLDPVVTNRKMLNWSDEDDGGSEITCFIIERKDAKMHTRWQ 8748
QY 179 ----- 178
Db 8749 PIETERSKCDITGLLEGOEYKFRVIKAKNKGCGPPVEIGPILAVDPLGPPTSPERLTYTE 8808
QY 179 ----- 178
Db 8809 RQRSTITLDWKEPRNGSGSPIQGYIIEKRHRDKPDERVNRKLCPTTSTFLVENLDEHQMY 8868
QY 179 ----- 178
Db 8869 EFRVAVNEIGESBPSLPLANVVIQDDEVPTTIKLRLSVRGPTIKVKAGEPVHIPADVTGL 8928
QY 179 ----- 178
Db 8929 PMPKLEWSKNETVIEKPTDALQITKEEYRSSEAKTELSIPKAVREDKGTIVTIVTASNRLGS 8988
QY 179 ----- 178

Db 8989 VFRNVHVEVDRPSPRNLAVTDIKABSCYLTDAPLDNGSGSEITHYVIDDKRDASRKA 9048
QY 179 ----- 178
Db 9049 WEEVTNTAVEKRYGIWKILPNGOYEFVRVAVNKYGISDECKSDKVVIQDPYRLPQPGPKP 9108
QY 179 -----PNV 181
Db 9109 KVLARTKSMVSVTPPLDNGGSPITGWLEKREBEGSPYWSRVSRAPIITKVLGKGVFN 9168
QY 182 PR----- 183
Db 9169 PRLLGVKYQFRAMAINAAGIGPPSEPSDEPVAGDPIFPFGPPSPCEVKDKTKSSISLGW 9228
QY 184 ----- 183
Db 9229 KPPAKGGSPIKGYIVEMQEEGTTDKVRVNEBPKLITTCVCVVPNLKELRYFRVKA 9288
QY 184 ----- 183
Db 9289 EAGESEPSDTTGBIPATDIOEEPEVFDIAGQDCLVCKAGSQIRIPAVIKGRPTPKSSWE 9348
QY 184 ----- 183
Db 9349 FDGKAKKAMKDVHDIPEDAQLETAENSSVIIIECKRSHTKYSITAKNKGQKTANCR 9408
QY 184 ----- 183
Db 9409 VKVMDVPKDLKVSIDITRGSCLSWKMPDDDDGDRIKGYVIBKRTIDGKAWTKVNPDC 9468
QY 184 ----- 183
Db 9469 GSTTFVVVDLLSEQYEFVRVRAENRFGIGPPVETIQRRTARDPIYPPDPPIKLKIGLITK 9528
QY 184 ----- 183
Db 9529 NTVHLSWKPKNDGSGPVTHYIVBECLAWDPTGTKEARQCNKRDVELOFTVEDLVEGG 9588
QY 184 ----- 183
Db 9589 EYEFVRVAVNAAGSVKSPSATVGPCDCQRPDMPSPIDLKBFMEVBEGTNVINIVAKIKGVPF 9648
QY 184 ----- 183
Db 9649 PTLTWFKAPKKKPDNKEPVLVDTHVKNLVVDDTCTLVIPOSRRSDTGLYTTITAVNNLGT 9708
QY 184 ----- 183
Db 9709 SKENRLNVLGRPGPPGVPIKFSVSADQMTLSWPPKDDGSGKITNYVIEKREANRKTW 9768
QY 184 ----- 183
Db 9769 HVSSEPKECTYTIPLKLEHGYVFRIMAQNKYIGEPDSEPETARNLFSVPGAPDKPTV 9828
QY 184 ----- 183
Db 9829 SSVTRNSMTVNWEEPEYDGGSPVTIGYWLKMDTTSKRWRKVRNDRPIKAMTLGVSKVTGL 9888
QY 184 ----- 183
Db 9889 IEGSDYQFRVAVINAAAGVPASLPDPATARDPIAPGPPPPKVTDTWTKSSADLEWSPL 9948
QY 184 ----- 183
Db 9949 KDGSKVTGYIIVEYKEBKEWEKGDKEVRGTVKLVTGLKBGAFYKFRVSAVNIAGIGE 10008
QY 184 ----- 183
Db 10009 PGEVTDVIEMKDLRSLPDLQDASVRDRIVVHAGGVIIRIIAYVSGKPPPTVTWNMERTL 10068
QY 184 ----- 183


```
Db 10069 PQEATIEATISSMWIKNCORSHQGVYSLAKNEAGERKKTIIIVDVLDPGPVGTPELA 10128
QY 184 ----- 183
Db 10129 HNLTNESCKLTWSPEDDGGSPITWYIEKRESDRRAWTPVYTVTRQNAVTQGLIQGA 10188
QY 184 -----WNTKLYG- 191
Db 10189 YFFRIAENSIGMPFVETSEALVIREPITVPERPEDLEVKEVTQNTVTLTNPPKYDGG 10248
QY 192 ----- 191
Db 10249 SEIINYVLESRLIGTEKFKHTNDNLLSRKTVTKGLEGDTYFVRVSAVINVGQKPSFC 10308
QY 192 ----- 191
Db 10309 TKPITCKDELAPPTLHLDPRDKLTIRVGEAFALTGRYSGKPKVSWFKDEADVLEDDRT 10368
QY 192 ----- 191
Db 10369 HIKTPATLALEKIKAKRSDSGKVCVVVENSTGSRKGFQCVNVVDHPGPPVPSFDEVT 10428
QY 192 ----- 191
Db 10429 KQYMWISWKPPLDDGSKITNYIIEKKEVGKDVMPVTSASAKTCKVSKLLEKGDYIFR 10488
QY 192 ----- 191
Db 10489 IHAENLYGSDPLVSDSMKAKDRFRVPDAPDQPIVTEVKDSALVTWKNPHDGGKPTNY 10548
QY 192 ----- 191
Db 10549 ILEKRETSKRWARVTKDPIHPYTKFRVPDLLEGQCFRVSARENEIGIGDPSPPKPVF 10608
QY 192 ----- 191
Db 10609 AKDPIAKSPPNPEAIDTTCNSVDLTWQPPRHGGSKILGYIYEYQKVGDEEWRANHT 10668
QY 192 ----- 191
Db 10669 PESCPETKYVTGLRDGQYKFRVLAVNAAGESOPAHYPEPVLVKORLEPPELILDANMA 10728
QY 192 -----PTKVNVD- 198
Db 10729 REQHIKVGDTLRLSAIIRKGVPPKVTWKEDRDAPTKARIDVTPVGSKLEIRNAAHEDGG 10788
QY 199 ----- 198
Db 10789 IYSLTVENPAGSKTVSVKVLVDKPGPPRDLEVSEIRKDSCVLTWKEPLDDGGSVITNYV 10848
QY 199 ----- 198
Db 10849 VERRDASAQWSPLSATSKKSHFAKHLNEGNYQLFRVAENQYGRGPFVETPKPIKALD 10908
QY 199 ----- 198
Db 10909 PLHPGPPKDLHHVDVDKTEVSLVWKNKDRDGGSPITGYLVEYQEGTQDWIKFKVTNL 10968
QY 199 -----SQIYFLGL- 207
Db 10969 ECWVTGLOQKTYRFRVKAENIVGLPDTTIPIECQKLVPPSVVELDVKLEGLVWKAG 11028
QY 208 ----- 207
Db 11029 TTVRPPIIRGVVPVTAKTWTDGSEIKTDEHYTVETDNFSSVLITIKNCLRRDTGEYQITV 11088
QY 208 ----- 207
Db 11089 SNAAGSKTVAVHLTVLDPGPPGTPGINILDTPEHMTISWQPPKDDGGSPVINIVEKQD 11148
QY 208 -----TALLRYAQRC- 219
Db 11149 TRKDTWGVVSSGSKTKLIKPHLOKGCBYVFRVRAENKIGVGPPLDSTPTVAKHKFSPPS 11208
```

```
QY 220 ----- 219
Db 11209 PPGKPVVTDITENAATVSWTLPKSDGGSPITGYTMERREVTGKWVRVNTKPIADLKPRVT 11268
QY 220 ----- 219
Db 11269 GLYEGNTYFVRVPAENLAGLSKPSSDPIKACRPIKPPGPPINPKLKDKSRETADLWVT 11328
QY 220 ----- 219
Db 11329 KPLSDGGSPILGYVVEQKPGTAQWNRINKDELIRQCAFRVPGLEGNEYFRICAANIV 11388
QY 220 ----- 219
Db 11389 GEGSPRELAESVIAKIDILHPPEVELDVTCDVITVRVQOTIRILARVKRPEPDITWKE 11448
QY 220 ----- 219
Db 11449 GKVLVREKRVLDLIQDLPRVELQIKEAVRADHGKYIISAKNSSGHAQGSAINVNLDRPGPC 11508
QY 220 ----- 219
Db 11509 QNLKAVTNVTKENCITISWENPLDNGSSEITNFIYVRKPNQKWSIVASDVTKRLIKANLL 11568
QY 220 ----- 219
Db 11569 ANNEYFRVCAENKVGVGPTIETKTPIILAINPIDRPGEPENLHIADKGTFFYLKWRRPD 11628
QY 220 ----- 219
Db 11629 YDGGSPNLSYHVERRLKSGDDWERVHKSGIKETHYMDRCVENQIYEFRTVQTKNEGGESD 11688
QY 220 ----- 219
Db 11689 WKTTEVVVKEDLQKPVLDLKLGLVLTAKGDTIRLEAGVKGKPPFVAVNTKDKDATDLT 11748
QY 220 ----- 219
Db 11749 RSPRVKIDTRADSKFSLTKAKRSDGGKYVVTATNTAGSFVAYATVNVLDKPGPVRLKI 11808
QY 220 ----- 219
Db 11809 VDVSSDRCTVCWDPEDDGGCEIQNYILEKETKMWSTYSATVLTFGTTVTRLIEGNE 11868
QY 220 ----- 219
Db 11869 YIFRVAENKIGTGPTTESKPVIAKTKYDKPGRDPPEPVTKVSKEEMTVVWNPPEYDGGK 11928
QY 220 ----- 219
Db 11929 SITGYFLEKKEKHSTRWVPVVKSAIPERRMKVQNLDPDHEYQFRVKAENEIGIGEPSLPS 11988
QY 220 ----- 219
Db 11989 RPVVAKDPIDPPGPTNFRVVDDTKHSITLGMGKPVYDGGAPIIGYVVMEMPKIADASPD 12048
QY 220 ----- 219
Db 12049 EGWKRCAAAQLVRKEFTVTSLDENQBYEFVCAQNVQIGORPAELKEAIKPKELIEPPE 12108
QY 220 ----- 219
Db 12109 IDLDASMRKLVIVRAGCPIRLFAIVRGRPAPKTVRKVIGIDNVVRKGVQVDLVDTMFLVI 12168
QY 220 ----- 219
Db 12169 PNSTRDPSGKYSLTFLVNPAGEKAVFVNVRLDTPGVSCLKVSDVTKTSCHVSWAPPEND 12228
QY 220 -----TH- 227
Db 12229 GGSQVTHYIEKREADRKTSTVTPEVKKTSFHVNLNLPVGNYYFRVAVNEYGPGVPTD 12288
```

QY 228 ----- 227
Db 12289 VPKPVLASDPLSEPPDKLEATEMTKNSATLAWLPPLRDGKAGIDGYIISYREBEQPAD 12348
QY 228 ----- 227
Db 12349 RWTEYSVVKDSLIVVTGLKEGKKYKFRVAARNAVGLPREAEGVYEAKEQLPPKILMP 12408
QY 228 ----- 227
Db 12409 EQITIKAGKKLRIEAAHVYKPHPTCKWKKGDEVTTSHLAVHKADSSILIIKDVTRKD 12468
QY 228 ----- 227
Db 12469 SGVSLTAENSSGTDQKIKVVVMDAPGPPQPFDISDIDADACSLSMWHIPLDGGSNIT 12528
QY 228 ----- 227
Db 12529 NYIVEKDVSRGDWVTALASVTKTSRVLKLIPOGEYIFRVAENRFGISEPLTSPKMA 12588
QY 228 ----- 227
Db 12589 QPFPVGPSEPKNARVTKVNKOCIFVANDRPDSGGSPIIIGYLIERKERNLLWVKANDTL 12648
QY 228 ----- 227
Db 12649 VRSTEYPCAGLVEGLEYSFRIYALNKAAGSSPPSKTEYVYVARMVDPDPGKPEVIDTKST 12708
QY 228 ----- 227
Db 12709 VSLIWARPKHDGSKIIGYVBEACKLPGDKWVRCNTAPHQIQOEYATATGLEKAQYOFR 12768
QY 228 ----- 227
Db 12769 AIARTAVNISPPSESDPVTIIAENVPPRIDLSVAMKSLITVYKAGTVCLDATVFGKMP 12828
QY 228 ----- 233
Db 12829 TVSWKKGDTLLKPAEGIRKMAQRNLCTLELFSVNRKSDGYTITAEENSSGSKSATIKLV 12888
QY 234 ----- 233
Db 12889 LDKPGPASVKINKMYSDRAMLSWEPLEDGGEITNYIVDKRETSRPNWAQVSATVPIT 12948
QY 234 ----- 233
Db 12949 SCSVEKLIEGHEYQFRI CAENKYGVDPVFTEPAIAKNPYDPPGRCDDPVISNITKDHT 13008
QY 234 ----- 233
Db 13009 VSWKPPADGGSPITGYLLEKRETOAVNWKVNRKPIIERTLKATGLQEGTEYEFRTAI 13068
QY 234 ----- 233
Db 13069 NKAGPKPSDASAAAYARDPQVPAPPAPPKYDYDTRSSVSLWGKPAYDGGSPIIGYLV 13128
QY 234 ----- 233
Db 13129 EVKRADSNVRCNLPQNLQKTRFEVTGLMBDTQYQFRVYAVNKIGYSDPSDVPDKHPK 13188
QY 234 ----- 233
Db 13189 DILIPPEGEHADLRKTLILRAGVTMRLYVPVKGRPPPKITWSKPNVNLDRDIGDIKST 13248
QY 234 ----- 233
Db 13249 DFDPLRCENNVKYGAKYIITLSENGCKEYTIIVKVLDPFGPPINVTVKEISKDSAYV 13308
QY 234 ----- 233
Db 13309 TWEPPIIDGGSPINVVYQKRAERKSWSTVTTECKSTSFVRPNLEBCKSYFFRPAENE 13368
QY 234 ----- 233

Db 13369 YGIGDPGETRDAVKASQTPGPPVDLKVRSVSKSSCSIGWKPHSGSGRIIGYVVDFLTE 13428
QY 234 ----- 233
Db 13429 ENKQVRVMSLSLOVSADKLTGEGKEYTFRVSAENENGEGTSEITVVARDWDVAPDLK 13488
QY 234 ----- 233
Db 13489 GLPDLCYLAKENSFRLKIPIKGKAPSVSWKGEDPLATDTRVSVSSAVNTTLIVYDC 13548
QY 234 ----- 233
Db 13549 QKSDAGKYTIITLKNVAGTKEGTISIKVVGKPGIPTGPIKPDVETAEAMTLKWA PKDDGG 13608
QY 234 ----- 233
Db 13609 SEITNYILEKRDSVNNKWVTCASAVQKTTFRVTRLHEGMEYTFRVSANENKYGVGGLKSE 13668
QY 234 ----- 233
Db 13669 PIVARHPFDVPDAPPPPNIVDVHRDSVSLTWDPKKTGGSPITGYHLEFKERNLLWKRA 13728
QY 234 ----- 236
Db 13729 NKTPIRMDFKVTGLTEGLEVEFRVMAINLAGVGKPSLPSEPVVALDPIDPPGKPEVINI 13788
QY 237 ----- 245
Db 13789 TRNSVTLIWTEPKYDGGHKLTYIIVEKRDLPFSKWMKANHVNPECAFTVTDLVEGGKYE 13848
QY 246 ----- 245
Db 13849 FRIRAKNTAGAI SAPSESTETIICKDEYEAPTVLDPTIKDGLTIKAGDTIVLNAISILG 13908
QY 246 ----- 245
Db 13909 KPLPKSSWSKAGDIRPSDITQITSTPTSSMLTIKYATRKDAGEYITITATNPFQTKVEHV 13968
QY 246 ----- 245
Db 13969 KVTVDVPGPPGVEISNVAEKATLTWTPLEDGSGSPIKSYILEKRETSRLLTWVSED 14028
QY 246 ----- K 246
Db 14029 IQSCHEVATKLIQNEVIIFRVSANVHYKGEVPQSEPVKMDRPGPPGPEKPEVSNVTK 14088
QY 247 NT-----MRKLK----- 253
Db 14089 NTATVSKRPVDDGGSEITGYHVREBKSLRWVRAIKTPVSDLRCKVTGLQEGSTYEF 14148
QY 254 ----- 253
Db 14149 VSAENRAGIGPPSEASDVLMOAAAYPPGPPSNPHVTDTTKKSASLAWGKPHYDGGLEIT 14208
QY 254 ----- 253
Db 14209 GYVVEHQVGEAWIKDITGTALRIITQFVVPDLQTKYKNFRISAINDAGVGEPAVIPDV 14268
QY 254 -----RKQAP----- 258
Db 14269 EIVEREMAPDFELDAELRLTVVRAGLSIRIFVPKGRPAPEVTWTKNINLKNRANIE 14328
QY 259 ----- 258
Db 14329 TESFTLLIIPECNRYDTGKFVMTIENPAGKSGFVNVVRLDTARPSQLRPTDITKDSVT 14388
QY 259 ----- 258
Db 14389 LHWDLPLIDGSRITNYIVEKEATRKSYSTATTKCHKCTYKVTGLSEGCEYFFRVAEN 14448
QY 259 ----- 258

Db 14449 EYGIGETETTPVKASEAPSPDNLMDITKSTVSLAWPKPKJGSGKITGYVIEAQR 14508
 QY 259 -----VKE----- 261
 Db 14509 KGSQWTHITTVKGLECVVRNLTEGEYTFQVMVNSAGRSAPRESRPVIVKEQTMPLPEL 14568
 QY 262 ----- 261
 Db 14569 DLRGYQKLVIAKAGDNIKVEIPVLRPKPTVTKGDOILKQTORVNFETTATSTILNI 14628
 QY 262 ----- 261
 Db 14629 NECVRSDSGPPLTARNIVGEVGDVITIQVHDIPOGPTGPDKFDEVSDFVTFPSWDPPEN 14688
 QY 262 -----QFEKA----- 267
 Db 14689 DGVPSINXVEMRQDSTTWVETLATTVIRTYKATRLTTGLEYOFRVKAQNRVGVGPGI 14748
 QY 268 ----- 267
 Db 14749 TSANIVANYPKVPGPGTPOVAVTKDSMTISWHEPLSDGSPILGYHVERKERNILW 14808
 QY 268 ----- 267
 Db 14809 QTVSKALVPGNIFKSSGLTDGIAYEFRIAENMACKSKPSKSEPMALDIPDPGKVPV 14868
 QY 268 ----- 267
 Db 14869 LNIIRHTVTWKWPEYTGPKITSYIVEKRDLPNGRWLKNFNSILENEFTVSLGTEA 14928
 QY 268 ----- 267
 Db 14929 AYEPRVIAKNAAGAISSPSPSDAITCRDDVEAPKIKVDVKFOTVILKAGEAFRLEADV 14988
 QY 268 ----- 267
 Db 14989 SGRPPPTMWSKDGKELEGTAKEIKIADFTNLVNKOSTRRDSGAYTLTATNPGGFANH 15048
 QY 268 ----- 267
 Db 15049 IPNVKVLDRPGPEGLAVTEVTSEKCVLSWFPPLDDGAKIDHVIQKRETSRLAWTNV 15108
 QY 268 ----- 267
 Db 15109 ASEQVQTKLVTKLLKGNIEYIFRVMANVKYGVGEPLSEPLAVNPYPPDPKPNPEVTT 15168
 QY 268 -----KKT----- 270
 Db 15169 ITKDSMVVWGHPSDSDGSGEIIINYIVERRDDKAGORWIKCNKKTLTDLRYKVSGLTEGHEY 15228
 QY 271 -----QSTTTPYF----- 278
 Db 15229 EFRIMAENAGISAPSPTSIFYKACDVFPGPGPNPRVLTSSRSISIANKNKIYDGS 15288
 QY 279 ----- 278
 Db 15289 EITGMVEIALPEDEWQIVPPAGLKATSYITGLTENQEKIRIYAMNSEGLGEPALV 15348
 QY 279 ----- 278
 Db 15349 PGTPKAEDRMPLPEIELDADLRKVVTIRACCTLRLFPVPIKGRPDPEVKWARDHGESLDKA 15408
 QY 279 -----SVT----- 281
 Db 15409 SIESASSYITLLIVGNVNRFDGSKYILTIVENSNGSKSAFVNVRLDTPGPPQDLKVEYTK 15468
 QY 282 ----- 281
 Db 15469 TSVTLTWDPLLDGSGKIKNIVIEKRESTRKAYSTVATNCHTSKWVDQLQEGCSYIFRV 15528
 QY 282 ----- 281
 Db 15529 LAENEYIGLPAETAESVKASERPLPPGKITLMDVTRNSVSLSWKPEHGGSRILGYIV 15588

QY 282 -----TSAALNVT----- 289
 Db 15589 EMQTKGSDKWATCATVKVTEATITGLIQGBEYSFRVSAQNBKIGSDPRQLSPVPIAKDLV 15648
 QY 290 ----- 289
 Db 15649 IPPAPKLLFNFTFTVLAGEDLKVDPFFIGRPTPAVTWTKDNVPLKQTTVRNAESTENNSLL 15708
 QY 290 ----- 289
 Db 15709 TIKDACREDVGHVVVVKLTNSAGEALETNLVILDKPGPTGPKWDEVTADSIITLSWGP 15768
 QY 290 ----- 289
 Db 15769 KYDGGSSINNYIVEKRDSTTTTWOIVSATVARTTIKACRLTKGCEYQFRIAAENRYGKST 15828
 QY 290 ----- 289
 Db 15829 YLNSEPTVAQYPPKVPGPPTVTLSSRDSMEVQMNPEISDGGSRVIGYHLERKERNIS 15888
 QY 290 ----- 289
 Db 15889 LNVKLNKTPIPQTKFKTTGLEEGVEYEFVSAENIVGIGKPSKYSECYVARDPCDPPGRP 15948
 QY 290 ----- 289
 Db 15949 EAIIVTRNSVTLQWKPTYDGGSKIYIVBEKELPEGRMKASFNTIIDTHFEVTLGVE 16008
 QY 290 ----- 289
 Db 16009 DHRYEPRVIAARNAAGVSESPSESTGAITARDEVDPPRISMDPKYKDTIIVHAGSFKYDA 16068
 QY 290 ----- 289
 Db 16069 DIYKPIPTQWIKGDOELNSTABLEIKSTDFAISLVKDAVRVDSGNVILKAKVAGER 16128
 QY 290 ----- 289
 Db 16129 SVTVNVKVLDRPGPGPVISGVTAECTLAWKPLQDGGSDIINYIVERRETSRLVMT 16188
 QY 290 ----- 289
 Db 16189 VVDANVQTLCKVTKLLEGNEYTFPRIMAVNKYGVGEPLSEPLSEPVVAKNPFVVPDAPKPEV 16248
 QY 290 ----- 289
 Db 16249 TTVTKDSMIVVWERPASDGGSEILGYVLEKDKBGIWTRCHKRLIGELRLRVTLGIENH 16308
 QY 290 ----- 289
 Db 16309 DYBFRVSAENAGLSEPPSPSAYQACDPIYKPGPPNNPKVIDITRSSVFLSWSKPIYDG 16368
 QY 290 ----- 289
 Db 16369 GCEIQYIVEKCDVNVGEMTCTPTTGINKTNIIEVEKLEKHEYNFRICAINKAGVGEHA 16428
 QY 290 ----- 289
 Db 16429 DVPGPPIIVEKLEAPDIDLLELRKIIINIRAGGSLRLFPVPIKGRPTPEVKWGVDEIRD 16488
 QY 290 ----- 289
 Db 16489 AAIIIDVTSSFTSLVDVNVRYDSKTYTLTLENSGTSKSAFVTVRLDTPSPVNLKVTEI 16548
 QY 290 ----- 289
 Db 16549 TKDSVSIWBPPLLDGGSKIKNIVIEKRETRKSYAAVVTNCHKNWKIDQLQEGCSYIF 16608
 QY 290 ----- 289
 Db 16609 RVTAEYIGLPAQTADPIKVAEVPQPGKITVDDVTRNSVSLSWTKPEHGGSKIIQY 16668

QY	290	-----TNVT-----	293
Db	16669	IVEMQAKHSEKWCARVKSLOAVITNLQGBEYLFRRVAVNEKGRSDPRSLAPIVAKD	16728
QY	294	-----	293
Db	16729	LVIEPDVKPAPSSYSVQVQDLKMEVPIISGRPKPTITWTGDLPLKQTTTRINVTDLSLDT	16788
QY	294	-----YSITTA-----	299
Db	16789	TLSEIKETHKDGQGYGIVANVVQKTASIEIVTLDKPPPKGPKVFPDDVSAESITLSMN	16848
QY	300	-----	299
Db	16849	PPLYTGGCQITNVIQKRDTTTTVMDVVSATVARTITLKTKTGTQYQFRIFAENRYCQ	16908
QY	300	-----	299
Db	16909	SPALESDPIVAQYPYKEFGPGTGFATAISKDSMVIQWHEPVNNGGSPVIGYHLERKERN	16968
QY	300	-----	299
Db	16969	SILNTKYNKTIIDHTQFRAQNLBEGIEYEFVRVAENIVGVGASKNSCYVARDPCDPPG	17028
QY	300	-----	299
Db	17029	TPEPIWKRNEITLQWTKPVYDGGSMITGYIVEKRDLPDGRWMKASFTNVIETQFTVSGL	17088
QY	300	-----	299
Db	17089	TEDQYEFVRVIAKNAAGAIKSPSDTGPIAKDEVELPRISMDPKFRDTIVVNAGETFRL	17148
QY	300	-----	299
Db	17149	EADVHGKPLPTIEWLRGDKIEESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNVAG	17208
QY	300	-----	299
Db	17209	SKSPFVNVKVLDRPGPEPVQVGTSEKCSLTWSPPLDGGSDISHYVVEKRETSRLA	17268
QY	300	-----	299
Db	17269	WTVVASEVVTNSLKVTKLLEGNEYVFRIMAVNKYGVGEPLSAPVLMKNPFLPPOPKSL	17328
QY	300	-----ARRV-----	303
Db	17329	EVTNIAKDSMTVCWNRPDSDGSEIIGYIVEKDRSGIRWIKCNKRITDRLRLVGTGLTE	17388
QY	304	-----STSTIAYR-----	311
Db	17389	DHEYEFVRSAENAGVGEPSPATVYKACDPVKPGPPTNAHIVDTTKNSITLAWGKPIY	17448
QY	312	-----	311
Db	17449	DGGSEILGYVVEICKADEEWOIVTPQGLRVTRFEISKLTQHOYKIRVCALNKVGLGE	17508
QY	312	-----	311
Db	17509	ATSVPGTVPKPEDKLEAPELDLDELKGIIVVRAGSARIHIPFKGRPMPEITWSREGEF	17568
QY	312	-----	311
Db	17569	TDKVQIEKGVNVTQLSIDNCDRNDAGKYILKLENSGSKSAFVTVKVLDTGPPQNLAVK	17628
QY	312	-----	311
Db	17629	EVRKDSAPLWEPPIIDGAKVKVNVVDKRESTRKAYANVSSKSKTSFKVENITEGAIY	17688
QY	312	-----	311
Db	17689	YFRVMAENEFGVGVVETVDVAKAAEPSPPGKVTLLTDSQTSASLWKEPEHDGGSRLV	17748
QY	312	-----	311

Db	17749	GYVVMQPKGTEKWSIVAESKVCNAVVTGLSSGOEYQFRVKAAYNEKSGKSDPRVLGVPVIA	17808
QY	312	-----	311
Db	17809	KDLTIQPSLKLFFNTYSIQAGEDLKIEIPVIGRPRPNISWVKDGBPLKQTTTRVNVVEETAT	17868
QY	312	-----	311
Db	17869	STVLHIKEGKDDFGKYTVTATNSAGTATENLSVILEKPGPPVGPVRFDSVADFVVIS	17928
QY	312	-----	311
Db	17929	WEPAYTGGCQISNYIVEKRDTTTTTWHMVSATVARTTIKITKLTGTQYQFRIFAENRY	17988
QY	312	-----	311
Db	17989	GKSAPLDSKAVIQYPFKEPGPGTGFVTSISKQMLQWHEPVNDGGTKIIGYHLEQKE	18048
QY	312	-----	311
Db	18049	KNSILWKLKNTPIQDTFKTTGLDLEGEYEFKVSANIVGIGKPSKVSECFVARDPCDP	18108
QY	312	-----PDSSPMKS-----	319
Db	18109	PGRPEAIVITRNNTVLAKWKPAYDGGSKIITGYIVEKKDLDPGRWMKASFTNVLETEFTVS	18168
QY	320	-----	319
Db	18169	GLVEDQYEFVRVIAARNAAGNFSEPSDSSGAIARDEIDAPNASLDPKYKDVIVVHAGETF	18228
QY	320	-----	319
Db	18229	VLEADIRGKIPDPVVWSKDCKELEETAARMEIKSTIOKTTILVVKDCIRTDGGQYILKLSN	18288
QY	320	-----	319
Db	18289	VGGTKSIPITVKVLDRPGSPGPKLVGTVAEKCYLAWNPLODGGANISHYIIEKRETS	18348
QY	320	-----IMA-----	322
Db	18349	RLSWQTVSTEQVQALNYKVTKLLPCNEYIFRVMVKNKYGIGEPLESGPVTACNPKPPGPP	18408
QY	323	-----TOLR-----	326
Db	18409	STPEVSALTKDSMVVTVARVPVDDGGTEIEGYILEKRDKEGVRTKCNKKTLDLRLVGT	18468
QY	327	-----	326
Db	18469	LTEGHSYEFVRVAABNAAGVGEPSPSVRYACDALYPPGPPSPKVTDTSRSSVSLAWSK	18528
QY	327	-----	326
Db	18529	PIYDGGAPKGYVVEVKEAAADEWTTCTPPTGLQKQFTVTKLKENTENFRICAINSEG	18588
QY	327	-----	326
Db	18589	VGEPATLPGSVVAQERIEPPEIELDADLRKVVVLRASATLRLFTVTKGRPEPEVKWEKAE	18648
QY	327	-----	326
Db	18649	GILTDRAQIEVTSFTMLVIDNVTRFDSGRYNLTLNENSGSKTAFVNVRLDSPSPAPVNL	18708
QY	327	-----	326
Db	18709	TIREVKKDSVTLSEWEPPLIDGGAKITNYIVEKRETRKAYATITNNTKTTFRLENLQEG	18768
QY	327	-----	326
Db	18769	CSYIFRVLASNEYGIGLPAETTEPVKVSBPPLPGPVRTLVDTVRNTATIKWEKPESDGGS	18828
QY	327	-----	326

Db	18829	KITGVVEMQTKGSEKMSCTQVTKLEATISGLTAGEEYVFRVAANVEKGRSDPRLGVP	18888
Qy	327	-----	326
Db	18889	VIARDIEIKPSVELPFTFNKAREQLKIDVPFKGRQATVNRKDGQTLKETTTRVNVSS	18948
Qy	327	-----	326
Db	18949	SKVTLSLKEASKEDVGTVELCVSNSAGSITVPTIIVLDRPGPGPIRIDEVSCDSIT	19008
Qy	327	-----	326
Db	19009	ISWNPPEYDGGCQISYIIVEKETTSTTHIVSQAVARTSIKIVLTTGSEYQFVCAEN	19068
Qy	327	-----	326
Db	19069	RYGSSYSESSAVAEYFSPPGPGTPKVHATKSTMLTWQVNVNDGSGRVIGHLEY	19128
Qy	327	-----	326
Db	19129	KERSILMSKANKILIAITQVKVSGLDBGLMYEYRYAENIAGICKSKCEPVPARDPC	19188
Qy	327	-----	326
Db	19189	DPPGQPEVNTIRKSVSLKSKPHYDGGAKITGYIVERRELDPDGRWLKCNNTNIQETFE	19248
Qy	327	-----	326
Db	19249	VTELTEDQRYEPRVFARNADSVSBSSESTGPIIVKDDVEPRVNMVKFRDVIUVKAGE	19308
Qy	327	-----	326
Db	19309	VLKINADIAGRPLPVISWAKGIEIEERARTEIISTDNHTLLTVKDCIRRDGTQVVLTK	19368
Qy	327	-----	326
Db	19369	NVAGTRSVAVNCKVLDPGPPAGPLEINGLTAEKCSLSWGRPOEDGGADIDYHKKRET	19428
Qy	327	-----	326
Db	19429	SHLAWTICEGELQMTCKVTLLKGNBYIFRVTGVNKGVGLEBSVAIKALDPPTVPSP	19488
Qy	327	-----	326
Db	19489	PTSLEITSVKESMTLCWRSPESDGSBISGYIIRERKNSLRVVRNKKPVDLRVKST	19548
Qy	327	-----	326
Db	19549	GLREGCEYRYAENAGLSLPSSETSPLIRABDPVFLPSPSPKIVDSGKTTITIAWV	19608
Qy	327	-----	326
Db	19609	KLPDPGGAPITGYTVEYKSDTDWKTISIQLRGTEYITISGLTTGAEYVFRVKSUNKVA	19668
Qy	327	-----	326
Db	19669	SDPSSDPQIAKEREEREPFLPIDSEMRKTLIVKAGASFTMTVPRGRPVNVLWSKPD	19728
Qy	327	DLAT-----	330
Db	19729	DLRTRAYVDTTDSRTSLTIENANRDSKTYTLTIQNLVLSAASLTLVVKVLOTPOGPTNIT	19788
Qy	331	-----	330
Db	19789	VQDVTKESAVLSWDVPENDGGAPVKNYHIEKREASKKAWSVTNNCNELSYKVTNLQEGA	19848
Qy	331	-----	330
Db	19849	IYYFRVSGENBFGVIGPAETKEGVKITEKSPPEKLGVTISKDSVSLTLKPEHDGGR	19908
Qy	331	-----	332
Db	19909	IVHYVVEALEKQKQNMVKCAVAKSTHVVVSGLRENSYFFRFFAENQAGLSDPRELLPV	19968
Qy	333	-----	332
Db	19969	LIKEQLEPEIDMKNPFSTHYVVRAGSNLKVDIIPISGKPLPKVTLSDRGVPLKATMRFT	20028
Qy	333	-----	332
Db	20029	EITAEMLTNLKBSTADAGRYEITANSSGTTKAFINIVVLDPRGPPTGPVVISDIITEE	20088
Qy	333	-----	332
Db	20089	SVTLKWBEPKYDGGSOVTNYILLKRETSATVMTVEVSATVARTMMKMKLTTGEBEYQPRIK	20148
Qy	333	-----	335
Db	20149	AENRFGISDHIDSACVTVKLPYTPPGPSTPWVNTVRESITVGMHEPVSNGSAVVGYH	20208
Qy	336	-----	335
Db	20209	LEMKDRNSILMQKANKLVIRTHFKVTTISAGLIYEFVRVYAENAGVGKPSHPSEVLAI	20268
Qy	336	-----	335
Db	20269	DACEPPRNVRTDISKNSVLSWQOPADGGSKITGYIVERREDLPDGRWTKASFNTVTE	20328
Qy	336	-----	339
Db	20329	QFTISGLTQNSQYEFVRVARNAGSISNPSEVVGPIITCIDSYGGPVIDLPLEYTEVVKYR	20388
Qy	340	-----	339
Db	20389	AGTSVKLRAGISGKPAFTIEWYKDDKELQTNALVCVENTTOLASILINKAORLNSGCYEL	20448
Qy	340	-----	339
Db	20449	KLRNAMESATIRVQILDKPGPGPIEFTVTAEKITLLWRPPADGGAKITHYIVBK	20508
Qy	340	-----	339
Db	20509	RETSRVVMSVSEHLEECIITTKIIGNEYIFRVRVAVNKYIGIGEPLESDSVAKNAFVT	20568
Qy	340	-----	339
Db	20569	PGPGPIPEVTKITKNSMTVVMRPIADGGSIDISGYLEKRDKSLGWFVKLKETIRDRQ	20628
Qy	340	-----	347
Db	20629	KVTGLTENSQYQYRVCANAGQGFSEPFYKAADPIDPPGPAKIRIADSTKSSITL	20688
Qy	348	-----	347
Db	20689	GWSKPVYDGGSAVTGYVVEIRQGBEEBWTTVSTKGEVRTTEYVYVNLKPGVNYVFRVSAV	20748
Qy	348	-----	347
Db	20749	NCAGGEPIEMNEPVOAKDILEAPEIDLVALRTSVIAKAGEDVQVLIPFKRPPPTVTM	20808
Qy	348	-----	351
Db	20809	RKDEKNLGSADARYSIENTDSSLLTIPQVTRNDTGKYLITTIENGVEGPKSTSVVKVLD	20868
Qy	352	-----	351
Db	20869	PAAQCLQVHVSRGTVTLLWDPPLIDGGSPIINNVIEKRDATAKRTWSVSHKCSSTSPK	20928
Qy	352	-----	351
Db	20929	LIDLSEKTPFPVLAENEIGIGEPCEPTEPVAEVPAPIRDLMSKDKSTKTSVILSWTK	20988
Qy	352	-----	351
Db	20989	PDFDGGSVITEYVVERKKGQWTHAGISKTCSIEVSQLEKQSVLEFVRVPAKNEKGLSD	21048

QY 352 ----- 351
 Db 21049 PVTIGPITVELIITPEVDLSIDPGAQVTVTRIGHNVHLELPYKPKPSISMLKDGPLK 21108
 QY 352 ----- 351
 Db 21109 ESEFVRFSKTENKITLSIKNAKKEGKVTVIDNAVCRIAPVITVITLGPSPKPGPIR 21168
 QY 352 ----- 351
 Db 21169 FDEIKADSVILSDVPEDNGGGEITCYSEKRETSQTNKMWCVSSVARTTFKVPNLVKDA 21228
 QY 352 ----- 351
 Db 21229 EYQFRVRAENRYGVSQPLVSSIIVAKHQPRIIPGPPGKPVYNNVTSNGMSLTWDAPVYDGG 21288
 QY 352 ----- 351
 Db 21289 SEVTGFHVEKKERNILMOKVNTSPISGREYRATGLVEGLDYQFRVYAENSAGLSSPDP 21348
 QY 352 ----- 351
 Db 21349 SKFTLAVSPVDPGTPDYIDVTRETITLKNPPLRDGSKIYGVYSEIKRQGNRWRNCF 21408
 QY 352 ----- 351
 Db 21409 TDVSECOQVTVTGLSPGDRYEFRIIARNAVGTISPSPSSGIIIMTRDENVPPIVEPGPEYF 21468
 QY 352 ----- 351
 Db 21469 DGLIISGESLRIKALVQGRVPRVTFWKGVEIEKRMMEITNVLGSLFVRDTRDH 21528
 QY 352 ----- 351
 Db 21529 RGVYTVKAGSASAKAEIKVKVQDTPGKVGPPIRFTNITGEKMTLWMDAPLNDGCAPIT 21588
 QY 352 ----- 356
 Db 21589 HYIEKRETSRLAWALIEDKCEAQSAITAKLINGNEYQFRVSAVNVKFGVGRPLSDPWA 21648
 QY 357 ----- 356
 Db 21649 QIQYTVDPAGIPEPSNITGNSITLTWARPSDGGSEIQOYILERREKSTRWVKVISKR 21708
 QY 357 ----- 356
 Db 21709 PISETRFKVTGLTEGNEYEFHMAENAAGVGPASGISRLIKREBPVNPDPPTVVKVYDT 21768
 QY 357 ----- 356
 Db 21769 SKTTVSLEWSKPVFDGMEIIGYIEMCKTDLGDWHKYNABACVKTRYTVDLQAGBEYK 21828
 QY 357 ----- 363
 Db 21829 FRVSAINGAGKDSCEVTGTIKAVDRDLTAPELDIDANFKQTHVVRAGASIRLFAYQGRP 21888
 QY 364 ----- 363
 Db 21889 TPTAWSKPDNLSLRADIIHTDGSFTUTVENCNNDAGKYTLTVENNSGKSITFTVKV 21948
 QY 364 ----- 363
 Db 21949 LDTGPGPGFITFKOVTRGSATLMWDAPLLDGGARIHHYVVEKREASRRSMQVISEKCTRQ 22008
 QY 364 ----- 363
 Db 22009 IFKVNDLAEGVYFRVSAVNEYGVEPEYEMPEIVATEQAPPRRLDVIDTSSSAVLA 22068
 QY 364 ----- 363
 Db 22069 WLKPDHIDGSRITGYLLEMROKQSDLMVEAGHTKQLTFTVERLVEKTEYEFVRVAKNDAG 22128
 QY 364 ----- 363

Db 22129 YSEPREAPSSVIIKEPQIEPTADLTGITNQLITCKAGSPFTIDVPISGRPAPKVTWKLEB 22188
 QY 364 ----- 363
 Db 22189 MRLKETDRVSITTTTKDRITLTKVDSMRGDSGRYFILTLENTAGVKTFSVTVVVVIGRPGPVT 22248
 QY 364 ----- 363
 Db 22249 GPIEVSSVSABSCVLSWMEPKDGGGTBITNYIVEKRESGTTAWQLVNSSVKRTQIKVTHL 22308
 QY 364 ----- 370
 Db 22309 TKYMEYFRVSSNENRFGVSKPLESAPIIAEHFPVPPSAPTRPEVYHVHSANAMSIRWEEPY 22368
 QY 371 ----- 370
 Db 22369 HDGSKIIGYVWEKKERNITLWVKENKVPCLCNKYKVTGLVEGLEQFRTVYALNAGVSK 22428
 QY 371 ----- 370
 Db 22429 ASEASRRPMAQNPDVAFGRPEVTDVTRSTVSLIWSAPAYDGGSKVVGYYIERKPVSEVGD 22488
 QY 371 ----- 370
 Db 22489 GRWLKCNVTIVSDNFFVTALSEGDTYEPFVLAKNAAGVISKGSESTGPTCRDEYAPPK 22548
 QY 371 ----- 370
 Db 22549 AELDARLHGLVTRIRAGSLVLDAAVGGKPEPKIITWTKGDKELDLCEKVSQYTGKRATA 22608
 QY 371 ----- 370
 Db 22609 VIKFCDRSDSGKYTLTVKNASGTVKAVSMVKVLDSPGPGKLTVSRVTQEKCTLWMSLPQ 22668
 QY 371 ----- 370
 Db 22669 EDGGAETHYIVERRETSRLNWIIVEGECPTLSVYVTRLIKNEYIFRVRVANKYGPVGP 22728
 QY 371 ----- 370
 Db 22729 VESEPIVARNSTIPSPGPIPEEVGTGKEHIIIQWTKPESDGGNEISNYLVDKREBSLR 22788
 QY 371 ----- 370
 Db 22789 WTRVNKYVYVDTLTKVTSLMEGCDYQFRVTAVNAAGNSEPSESNFISCREFSYTPGPP 22848
 QY 371 ----- 370
 Db 22849 SAPRVVDTTKHSISLAWTKPMYDGGTDIVGVYLEMOBKTDQWVRVHTNATIRNTEFTVP 22908
 QY 371 ----- 370
 Db 22909 DLKMGOKYFRVAANVKNMGSEYSEIAETEPVERIEIPDLELADDLKKTVTIRAGASLR 22968
 QY 371 ----- 370
 Db 22969 LMVSVSRPPPVITWSKQIDLASRAIIDTTESYSLLIYDKVNRDAGKYTIEAENOSGK 23028
 QY 371 ----- 370
 Db 23029 KSATVLVKVYDTPGPCFSVKVEVSRDSVTITWEIPTIDGAPINNVIVEKREAAWRAFK 23088
 QY 371 ----- 374
 Db 23089 TVTTCKSKTLYRISGLVEGTMHYFRVLPENIYIGEPCTSDAULVSEVPLVPAKLEVD 23148
 QY 375 ----- 374
 Db 23149 VTKSTVTLAWEKPLYDGSRLTGYLEACKAGTERMKVVTLLKPTVLEHTVTSLEGEQY 23208
 QY 375 ----- 381
 TLDMSL
 :|:|:|

Db 23209 LPRIARQNEKGVSEPRETTVAITVQDLRLVPLTDIDSLTSTPQKTIHVPAGRPVELVPIPIAGR 23268
QY 382 ----- 381
Db 23269 PPPAASWFFAGSKLRESERVTVETHTKVAKLFTIRETTTIDTGEYTLKLVNVTGTTSETIK 23328
QY 382 ----- 381
Db 23329 VIILDKPGPTGPIKIDEIDATSIITISWEPPELDGGAPLSGVYVBEORDAHRGWLVPVSES 23388
QY 382 ----- 381
Db 23389 VTRSTFKTRLTEGNEYVFRVAATNRFGLGSYLQSEVIECRSSIRIPGPPELQIFDVSR 23448
QY 382 ----- 381
Db 23449 DGMILTWPBDDGSGQVTGYIVERKEVRADRWVRNVKVPVTMTRYRSTGLTEGLEYEHR 23508
QY 382 ----- 381
Db 23509 VTAINARGSKPSRPSKPIVAMDPIAPPCKPQNPRVTDTRTSVSLAWSVPEDEGSKVT 23568
QY 382 ----- 381
Db 23569 GYLIEQKVQDQHEWTKCNTPTTKIRBYTLTHLPQGAERYFRVLACNAGCGPEPAEVPCTV 23628
QY 382 ----- 389
Db 23629 KVTEMLPYDPVELDERYQEGIFVRQGVIRLTIPIKGPFPICKWTKGQDISKEAMIAI 23688
QY 390 ----- 392
Db 23689 SETHTELVIKEADRGDSGTDLVLENKCGKKAUYIKRVIGSPNSPEGLBYDDIQVRSV 23748
QY 393 ----- 392
Db 23749 RVSWRPADGGADILGILERREVPAKAAWTTIDSRVRGTSLLVVKGLKENVEYHFRVSAE 23808
QY 393 ----- 392
Db 23809 NQFISGLPKSEEPVPTPLNPPPPNPPEVLDVTKSSVLSWSRPKDGGSRVTGY 23868
QY 393 ----- 406
Db 23869 IERKETSTDKVRHNKTOITMTYTVGLVPDAEYQFRIIAQNVDVGLSETSPASEPVVCK 23928
QY 407 ----- 409
Db 23929 DPFKPSQPGLEILSISKDSVTLQWKEPCDGGKEILGYVWVEYRQSGDSAWKSNKRI 23988
QY 410 ----- 409
Db 23989 KDKQFTIGLLEATEYFRVFAENETGLSRPRRTAMSIKTLTSGEAPGIRKEMKDVTK 24048
QY 410 ----- 409
Db 24049 LGAAQLSCQIVGRPLDIKWYRFGKELIQSRKYNMSDGRTHLTVMTEBEDEGVYTC 24108
QY 410 ----- 409
Db 24109 IATNEVGEVETSKLLQATQFHPGYPLKBYKYGAVGSTLRLHVMYIGRPVPAMTWPHG 24168
QY 410 ----- 409
Db 24169 QKLLQNSENITIENTEYTHLVKMNVRQKTHAGKYKVQLSNVFGTVDAILDVEIQDKPDK 24228
QY 410 ----- 409
Db 24229 PTGPVIEALLKNSAVISWKPADDGGSWITNYVVEKCEAEGAEWQLVSSAISVTTCRI 24288
QY 410 ----- 417
Db 24289 VNLTENAGYFRVSAQNTFGISDPLEVSSVVIKSPFEKPGAPGKPTITAVTKDSCVVAW 24348

QY 418 ----- 417
Db 24349 KPPASDGAIRNYYLEKREKKQNKWISVTTTEIRETVFSVKNLIEGLEYEYPRVKCNILG 24408
QY 418 ----- 417
Db 24409 GESEWSEISEPITPKSDVPIQAPHPKBEELNLRVYQSNATLVCKVTHPKPIVVKYRQG 24468
QY 418 ----- 417
Db 24469 KEIIADGLKYRIQBFKGGYHQLIIASVTDDDATVYQVRATNQGSGSVSGTASLEVEVPKI 24528
QY 418 ----- 417
Db 24529 HLPKTLGMAVGHALRGEVVISIKIPFGKPDVITWQKQDLIDNNHGYQVIVTRSF7SL 24588
QY 418 ----- 417
Db 24589 VFPNGVERKDAGFYVVCANRFGIDQKTVELDVADVPDPGRGVKVDASRDSVNIWTPEP 24648
QY 418 ----- 417
Db 24649 ASDGSKITNYIVEKCATTAERLVRGQARETRVTVINLPGKTSYQFPRVIAENKFGLSKP 24708
QY 418 ----- 417
Db 24709 SEPSEPTITKEDKTRAMNYDEVBDETRVSMTKASHSTKELYEKYMIAEDLGRCEGIV 24768
QY 418 ----- 417
Db 24769 HRCVETSSKTYMAKFKVKGTDQVLVKKEISILNIAHRNILHLHESFESMEELVMIPE 24828
QY 418 ----- 417
Db 24829 FISGLDIFERINTSAPELNEREIVSVYHVQCEALQFLHSHNIGHDFIRPENIIYQTRSS 24888
QY 418 ----- 418
Db 24889 TIKIIFQARQLKPGDNFRLLFTAPEYYAVEVHQHDVVSTATDMWSLGLTVYVLLSGIN 24948
QY 419 ----- 425
Db 24949 PFLAETNQIIEINMNAEYTFDEAPKESIEANMDFVDRLLVKKERKSRMTASEALQHPWL 25008
QY 426 ----- 425
Db 25009 KQKIERVSTKVIRTLKHRRYHTLIKKDLNMMVSAARISCGAIRSQKGVSAKVKVASI 25068
QY 426 ----- 433
Db 25069 EIGPVSGQIMHVGEGGHVKYCKIENYDQSTQVTWYFGVQLENSKEYEITVEDGVAI 25128
QY 434 ----- 433
Db 25129 LYVKDITKLDGTYRCVNDYDGEDSSYAEFLVKGVREVDYCRRTMKIKRRTDTMRL 25188
QY 434 ----- 433
Db 25189 LERPPEFTPLYNKTAUVGENVRFGVITVHPHPTHVWYKSGQIKPGDNKKYTFESDK 25248
QY 434 ----- 433
Db 25249 GLYQLTINSVTTDDAEYTVVARNKYGEDSCAKLTVTLHPPTDSTLPRMPFKRLANAE 25308
QY 434 ----- 433
Db 25309 CQEQSVCFEIRVSGIPPPPLKWEKQGPLSLGPNIEIIEHGLDYALHIRTLPEDTGY 25368
QY 434 ----- 433
Db 25369 YRVATNTAGSTSCQAHQVRLRYKKOEFKSKEHERHVQKIDKTLRMAEILSGTESV 25428

Qy 434 ----- 433
Db 25429 PLTQVAKALREAAVLKPAVSTKTVKGFRLEIEBKEERKLMPYDVPFRKYQTTI 25488
Qy 434 ----- 433
Db 25489 EBDQRIKQFVPM5DMKYKIRDOVEMPGKLDREVQKPRKRLRLSRWEOFYVMPPLRITD 25548
Qy 434 ----- 433
Db 25549 QYRPKWRIPKLSQDLDLEIVRPARRTPSPDYFYFRRRRSLGDISDELLLPIDDDYLAM 25608
Qy 434 -----SLR----- 436
Db 25609 KRTEBERLRLEBELELGF5SAGPPSRPPHFLSSRLYSSPOAHVKVETZRNKFRSTYTHI 25668
Qy 437 ----- 436
Db 25669 PTKAESTSYAELRRHQAAYRQPKQRQIRMAREDEELLRPVTTTQHLSEYKSELDFM 25728
Qy 437 -----SPTVNL----- 443
Db 25729 SKEEKSRKRRQREVTEITEIEYEIEISKAQRESSSASRLRRRRSLSPYIELMRP 25788
Qy 444 -----TP----- 445
Db 25789 VSELIRSPQAEYEDDTERSPTRPRSPSPVSSRSLSRPFSGARDFISRYES 25848
Qy 446 ----- 445
Db 25849 MKAALKTKTSEKYEVL5QOPFLDHPRIPLMRSHRVPQGNTRFILNVQSKPTARV 25908
Qy 446 ----- 445
Db 25909 KWHNGVELOESSKIHYNTSGVLTLLEILDHDDSGTYRAVCTNYKGEASDYALDVTG 25968
Qy 446 ----- 445
Db 25969 GDYTYASQRDEBVR5PPELTRTEAVANVP5FKTSEMEASSVREYK5QMTETRESL 26028
Qy 446 ----- 445
Db 26029 SSYHSASABMSAAL5EKS5TRKIKTTLAARILTKPR5MTVVEG5SAR5P5C5DTD 26088
Qy 446 -----PEHRAV----- 445
Db 26089 GEPVPTVWLKQVLST5ARHQVTTTKYKSTFEISSVQASDEGNYSVVVENSEGKQEA5 26148
Qy 446 -----PEHRAV----- 452
Db 26149 FTLTIQARVTEKAVT5PPRVK5PEPRVKSPEAVK5P5R5HP5KAV5PTETKPTP 26208
Qy 453 ----- 452
Db 26209 IEKVQHL5V5APPK1TQFLKAB5KEIAK1TCV5ESSVLRAKEVTWKDGK1KENGHFQ 26268
Qy 453 -----NLSTN----- 458
Db 26269 PHYSADGTVELKINNLT5DQGEYVEISGEGGTSKTNLQPMQAFK5IHEK5K1SETK 26328
Qy 459 -----SLW 461
Db 26329 K5DQKTT5TVTRKTEPKAP5PIS5KPV1VTGLQD5TVSSD5VAK5AVKATG5PRTAIW 26388
Qy 462 ----- 461
Db 26389 TKDGKAITQGGKYL5EDKGGFFLEIHK5TDSGLYTCVTKNSAG5SVSSCKLTIKAIK 26448
Qy 462 ----- 461
Db 26449 DTEAQK5V5TK5BITPQKAVQ5EISQKALR5EEIKM5EAK5Q5EKLAK5EASKVL5S 26508
Qy 462 ----- 461

Db 26509 EEVKK5AAT5L5K5IVH5EITK5Q5A5EEVTRH5AIKAF5Q5MSIN5EQRLVLKANIAGA 26568
Qy 462 -----W----- 462
Db 26569 TDVKWVLNGVELTNE5EYRV5G5GSDQTLTIKQASHRDEGILTCISK5KEGIVKQCYDLT 26628
Qy 463 -----WLQ 465
Db 26629 LSKELSDAPAFISQPR5QNINEQ5NVLFTR5EISG5P5EIEW5FK 26672
RESULT 3
S41309
C:Species: cyclosporin fungus
C:Species: Tolypocladium inflatum (cyclosporin fungus)
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C:Accession: S45487; S41309
R:Weber, G.; Schoergendorfer, K.; Schneider-Scherzer, E.; Leitner, E.
Curr. Genet. 26, 120-125, 1994
A:Title: The peptide synthetase catalyzing cyclosporine production in Tolypocladium nivei
A:Reference number: S45487; MUID:95094306; PMID:8001164
A:Accession: S45487
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-15281 <WEB>
A:Cross-references: UNIPROT:Q09164; EMBL:Z28383; NID:g440168; PIDN:CAA82227.1; PID:g44401
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
A:Note: only a part of the translation is shown
A:Note: the source is designated as Tolypocladium inflatum
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: cyclosporin fungus cyclosporin synthetase; acetate-CoA ligase homology; a
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:532-1008/Domain: acetate-CoA ligase homology <ACLI>
F:1029-1096/Domain: acyl carrier protein homology <ACLI>
F:1618-2069/Domain: acetate-CoA ligase homology <ACP1>
F:2527-2594/Domain: acyl carrier protein homology <ACP2>
F:3115-3553/Domain: acetate-CoA ligase homology <ACLI>
F:4014-4081/Domain: acyl carrier protein homology <ACP3>
F:4601-5057/Domain: acetate-CoA ligase homology <ACLI>
F:5506-5573/Domain: acyl carrier protein homology <ACP4>
F:6094-6546/Domain: acetate-CoA ligase homology <ACLI>
F:7003-7070/Domain: acyl carrier protein homology <ACP5>
F:7591-8042/Domain: acetate-CoA ligase homology <ACLI>
F:8063-8130/Domain: acyl carrier protein homology <ACP6>
F:8552-9157/Domain: acetate-CoA ligase homology <ACLI>
F:9558-9625/Domain: acyl carrier protein homology <ACP7>
F:10146-10586/Domain: acetate-CoA ligase homology <ACLI>
F:11055-11122/Domain: acyl carrier protein homology <ACP8>
F:11635-12106/Domain: acetate-CoA ligase homology <ACLI>
F:12127-12194/Domain: acyl carrier protein homology <ACP9>
F:12715-13159/Domain: acetate-CoA ligase homology <ACP9>
F:13623-13690/Domain: acyl carrier protein homology <ACP9>
F:14213-14676/Domain: acetate-CoA ligase homology <ACP10>
F:14698-14765/Domain: acyl carrier protein homology <ACP11>
F:1060,2558,4045,5537,7034,8094,9589,11086,12158,14729/Binding site: phosphopantet:

Query Match 23.5%; Score 575; DB 2; Length 15281;
Best Local Similarity 1.7%; Pred. No. 0.047;
Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;
Qy 1 NG----- 2
Db 1 MGAIGQDMAYDRLANP5R5SSIS5N5RY5EP5VQ5FAQ5RL5FLH5QLK5GAS5DIT5PAAIR 60
Qy 3 ----- 2
Db 61 LRGLDITDALNA5RALTQ5RHETLRTFT5KEQ5GVQ5V5VHAS5GL5R5L5RIV5DASS5RDLAQ 120
Qy 3 ----- 2
Db 121 LLABEQTMKFDL5SEPA5RW5ALLK5VAED5HHI5L5IVV5HHI5SD5R5L5DIT5QO5EL5ELYTAA 180

Qy 3 ----- 2
Db 181 SQKSIACPLGPIPIQYRDLTTWQNOQDEQVAOERQLGWYIEQLDNNTPAELLTELPRP 240
Qy 3 ----- 8
Db 241 AIPSGETKISFOIDGSHKELLAPCRSQQVTAVALLAAPRAVHFRITGAEDATIGAPV 300
Qy 9 ----- 12
Db 301 ANRDPELENNVAPLATIQCMRVVLDEDDTTESVLRQIMSVWTEAHANROVPPERIVSAL 360
Qy 13 ----- 17
Db 361 LPGSTDSRHLPLVQLMFALHPAQDTGRARWGFLAEATLQSAAPTRFDMEMHLPBGDDRFD 420
Qy 18 ----- 17
Db 421 ANVLFTGLFDAEAIKRSVVSIPREVLRRGISEPAVHVKTMTPLTDGLAAIRDMLLDIGTT 480
Qy 18 ----- 17
Db 481 DYPREASVDMFQEQVALNPATADASSRLSYSELDHKSQDLAAMLRRLQKPELTIG 540
Qy 18 ----- 17
Db 541 VLSPPSCETMVSFLGILKAHLAYLPDINVPLARIESILSAVDGHLVLLGSLNVPQKVD 600
Qy 18 ----- 17
Db 601 VPDVELLRISDALNGSQVGLAGQKATAPKPSATDLAYVIFTSGTGPKGVMIEHRGIVR 660
Qy 18 ----- 17
Db 661 LVKGTNIIIPAQAAPTAHLANIAFDLSTWEIYTPILNGGTLVCIHVSITLDSKALEAVP 720
Qy 18 ----- 23
Db 721 TKEGIRVAFLAPALIKOCLADRPALFAGLSLYAIGDRFDRRDALHAKSLVKGWYNAYG 780
Qy 24 ----- 23
Db 781 PTENSVVSTIYVSEASPFVTGVPVGRAISNGAYVMDQOQLVSPGVGELVVSQGLA 840
Qy 24 ----- 23
Db 841 RGYTDSALDKNRFVVVQIDGESIRGYRTGDRARYSLKGGQIEFFGRMDQVKIRGHRIE 900
Qy 24 ----- 23
Db 901 AVEHALLNSQVRDAVWIRRQEEBPAMIAFTVTTQTLDPHLVNINGHVPDGNQSK 960
Qy 24 ----- 23
Db 961 NDQPAVHVESLRRRLQMLLPSYMMPARIVVLHPLPNPKGVDRKALGOSAKTVQSKL 1020
Qy 24 ----- 23
Db 1021 VSQVAPRNEIEAVLCEEYRSVLGVVEGVTDNFPDLGHSITAMKLAARISQRLDIQASV 1080
Qy 24 ----- 23
Db 1081 ATVPEQPLADLAATIQRGSTLYSIVPTTEYTGPEQSPAQGLWFLQNLGASWYNVM 1140
Qy 24 ----- 23
Db 1141 LTVRLRGLDVLGALTALLALEKRHETLRTTFEERDGVGMQVHSSLMGELRLIDISEKS 1200
Qy 24 ----- 23
Db 1201 GTAAHEALMKEQSTRFDLTREPGWRVALLKLADHHIFSIVMHHIVSDGWSLDDLRLHELQ 1260

Qy 24 ----- 23
Db 1261 LYSAAALRGQDPLSRLEPLPIQYRDPFVWQKDSQKAAHQRLEYMTKQLADSTPAELLT 1320
Qy 24 ----- 23
Db 1321 DFPSPSILSGKAGKVPVPAIBESLYDTLQVFSRTHQVTSFAVLLAAFRAAHFRLTGSDNAT 1380
Qy 24 ----- 23
Db 1381 IGVPSANRPELENVIGFFVNTQCIRITIDENDFESLVRQVSRSTTTAAQNDQVPPFEQ 1440
Qy 24 ----- 23
Db 1441 VSSSLMPSSSRDASRNPLVQLMFALHGQDLFKIQLEGTSEBEVPTBEVTRFDIEPHLYQ 1500
Qy 24 ----- 23
Db 1501 GASKLSGDIIPAADLPEAETIRGVSVFQEVLRGLQOPQTPIMTMTPLTDGIPELERMGL 1560
Qy 24 ----- 23
Db 1561 LHMVKTDYPRNMSVVDVFPQQVRLSABATAVIDSSSRMSYAELOQRSQVAAWLRQRQLP 1620
Qy 24 ----- 23
Db 1621 AETFAVLAPRSCBAVIALFGILKAGHAYLPDVNVPAAURLRAYLAEVKGBKVLGAGE 1680
Qy 24 ----- 23
Db 1681 PSPEGQSPESIVRIADATSPAGHASLRDGSKSKTAGSLAVIFTSGTGPKGVMIEHR 1740
Qy 24 ----- 23
Db 1741 GVLRLVKQTNILSSLPPAQTFRMAHMSNLAFDASIVWEVFTALLNGGSLVCIDRFTILDAQ 1800
Qy 24 ----- 23
Db 1801 ALEALFUREHINTALPPPALLKQCLTDAATIKSLDLLLYVGGDRDLTADAAALAKALKSE 1860
Qy 24 ----- 23
Db 1861 VYNAYGTENTVMSTLYSIADTERFVNGVPIGRAVSNVGVVMDQNLQVLGVNGELVV 1920
Qy 24 ----- 23
Db 1921 TGDCLARGYTNPALDSRDFVDVVIARGQLLRAYRTGDRARYRPKQGVQVEFFGRMDHQVKVR 1980
Qy 24 ----- 28
Db 1981 GHRIELAEVEHALLSAGVHDVAVVSNQSDNQGVEMVAFITAQDNETLQEAQSSNQVOE 2040
Qy 29 ----- 28
Db 2041 WESHFTTAYADITADONTLGRDFTSWTSWYDGLIDKREMQEWLDDTMTFTLDGQAAG 2100
Qy 29 ----- 28
Db 2101 HVLEIGTGTGMVLFNLQAGLKSXYGLEPSQSAQVFNKAAQTPFGLGKAQVHVGTAMD 2160
Qy 29 ----- 28
Db 2161 TGRLSALSPDLIVINSVAQVPPSREYLAEEVVEALVRIPGVRRIFFGDMRTVATHKDFLVA 2220
Qy 29 ----- 28
Db 2221 RAVHTNGSKVTRSKVQOEVARLEBELELVDPAFFTSLSKESLESEEIHEVILPNMKVN 2280
Qy 29 ----- 28
Db 2281 NELSSRYGAVLHTRHNQNSRSHKINAESWIDPASSQMDROGLARLLKENKDAESIA 2340
Qy 29 ----- 28

Db	2341	VFNIPVSKTIVERHIAKSLADDDGDDTHSSIDGVAWISAAREKASQCPSLDVHDLVQLA	2400
Qy	29	-----	28
Db	2401	EDAGPRVEVSWARQSONGALDVFFHFPQPTENESRALVDFPTDYKGOQARSLTNRLQR	2460
Qy	29	-----	28
Db	2461	VESRRIEAQVREQLVLLPAYMIPARIIVVLQNMPLNTSGKVDKRLTLRAKTAARTPSS	2520
Qy	29	-----	NCKVMSKA 36
Db	2521	ELVAPRDSIEALICKFKDVLGVEVGI TDNFFNVGHSILATKLAARLSROLNAQIAVKD	2580
Qy	37	LYNRP-----	41
Db	2581	IFDRPVIADLAATIOODTTEHNPILPTSYTGPVEQSPAQRMLFQLDQLNVGATWYLMPPA	2640
Qy	42	-----	41
Db	2641	VRLRGPLVVSALAAALLAEERHETLRTTFIEQEGIMQVTHPPAPKELRVIDVSGEES	2700
Qy	42	-----	41
Db	2701	TIQKILEKEQTTPFNLAGPGRFRLAKKTGEDEHILSTVMHHAISDGWSVDIFQOEIGQF	2760
Qy	42	-----WR-----	43
Db	2761	YSAILRGHDPLAQIAPLSIQVRDFATWQRIQFQVAEHRRLQAYWTKQLADKNKPAELLTDF	2820
Qy	44	-----GLVLSK-----	49
Db	2821	KRPPLMSGRAGEIPVVVDGLIYEKLQDFCRINQVTAFTVLLAAFPAAHYRMGTGEDATIG	2880
Qy	50	-----	49
Db	2881	TPIANRNRPELEGLIGFVNTQCMRITVDVEDSFETLVHQVRETTAAAHANQDVPFEQIV	2940
Qy	50	-----IGKYKLDQKLEILLRQLETT-----	69
Db	2941	SNILPGSSDTSRNLVLQMLFALHSQONLGKVRLEGIEBIIISIAETTRFDIEFHLVQBAE	3000
Qy	70	-----	69
Db	3001	RLNGSIVVAADLFVPETIQSVITIFQGILOKGLGEPDMPVASMALDGGLESIRSTGLLHP	3060
Qy	70	-----	69
Db	3061	QOODYPCDASVVOIFKQOVVNPVDVIAVRDESTRLSYADLDRKSDQVACWLSRRGIAPET	3120
Qy	70	-----	69
Db	3121	FVAILAPRSCETIVAILGLVKANLAYLPDVNVPASRLAEILSEVSGSMLVLVGAETPIP	3180
Qy	70	-----	69
Db	3181	EGMAEATIRITEILADAKTDDINGLAASQPTAASLAVIIFTSGSTGRPKGVVMEHRGIV	3240
Qy	70	-----	69
Db	3241	RLTKQTNITSKLPSFHHMAHISNLAFDASVWEVFTLLNGGTLCIDYFTLLLESTALEKV	3300
Qy	70	-----	69
Db	3301	FFDQVNVALLPPALLKQCLDNPALVKTLVLYIGGDRLDASDAKARGLVQTAQFNAY	3360
Qy	70	-----ISTKYNVSKOP-----	80
Db	3361	GPTENTVMSTIYPIAEDPPFINGVPIGHAVSNSGAFVMDQNOQITPPGAGBELIVTGDGLA	3420
Qy	81	-----	80
Db	3421	RGYTTSSLNTGRFINVDIDGQVRAVRTGDRVVRPKDLQIEFFGRIDHQVKIRGHRIEP	3480
Qy	81	-----	80
Db	3481	AEVEYALLSHDLVTDAAVVVTHSQENQDLEMVGFVAARVADVREDESSNQVQWQTHFDSI	3540
Qy	81	-----	80
Db	3541	AYADITTTDQOSLGRDPMSTWMTDGSLLIKKSQOEWLDDTWRSLLDSQPPGHVLEVTG	3600
Qy	81	-----	80
Db	3601	TGMVLENLREGGLQSVVGLSPSATAFVNKAAKSPGLEDRIRVEVGTATDIDRLGDD	3660
Qy	81	-----VKMLT-----	85
Db	3661	LHAGLVVNSVAQYFPQDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDLVARAVHAL	3720
Qy	86	-----	85
Db	3721	GDKATKAEIQREVVRMESEDELLVDPAFFTSLTTOVENIKHVEILPKMRATNELSSYR	3780
Qy	86	-----	85
Db	3781	YAAVLHVNDLAKPAHKVSPGAWVDFAAATKMDRDALIRLLRGTKISDHIAIANIPNSKTIV	3840
Qy	86	-----	85
Db	3841	ERTICESVYDLGGDAKSDNRVSWLSAARNAVKVASLSAIDLVDIAQEAAGRVEISCAR	3900
Qy	86	-----	85
Db	3901	QWSQNGALDAVFHHLGPSQSSHVLIDFLTDHQRPEALTNHPLHRAQSRVVERQIRER	3960
Qy	86	MNTEFPQYIILA-----	97
Db	3961	LQTLTPAYMIPAQIMWLDKPLNANGKVDKRLQTORAQTPKAKQVAPVAPRTEIERVL	4020
Qy	98	-----	97
Db	4021	CQEPSVLGVNDIGIMENFFDLGGHSLMATKLAARISRRLETHSVSKSIFDHPRVCDLVI	4080
Qy	98	-----GPI-QNYSITYLMF-----	110
Db	4081	VQOGSAPHDPVIVSKYTKTGPVPQSPAQGRMLFDQLNFGATWYLMPLAVLRGAMNVHALT	4140
Qy	111	-----	110
Db	4141	AALLAERRHELLARTTTFYEQNGVMQKVPVVTETLRIIDLNSGDGDYDLPCLKKEQTAPF	4200
Qy	111	-----DFYSTOL-----	117
Db	4201	HLETEPCWRVALLRLPGDYILSVMMHIIISDGSVDVLFQELGQFYSTAVKGHDPLSQT	4260
Qy	118	-----RKP-----	120
Db	4261	TPLPIHYRDPALWQKPTQSEHERQLQYVWEQLVDSAPABELLTDLRPSILSQAGEMS	4320
Qy	121	-----AKY-----	123
Db	4321	VTIEGALYKNLEBFRCVRHVTSFVLLAALRAAHYRLTGSEDATIGTPIANRNRPELEQI	4380
Qy	124	-----VYSQYN-----	123
Db	4381	IGPFVNTQCIIRITVNEDETFESLVQVQRSTATAFAHQDVPFKEIVSTLLPGSRDASRNP	4440
Qy	124	-----VYSQYN-----	129
Db	4441	LVQLMFAVHSQKNLGEKLENAHSEVVPTBITTRFDLEPHLFQODDKLEGSILYSTDLFE	4500
Qy	130	-----	129
Db	4501	AVSQSLLSVFQBIILRRGLNGPDVPISLTLQDGIVDLQRLDQKLEVQKTYPRDSSVVDV	4560

QY 130 ----- 129
Db 4561 PHEQVINPDSIALIHGSEKLSYAQLDRESDRVARLHRFSFSDTLIAVLAPRSCETII 4620
QY 130 ----- 129
Db 4621 AFLGILKANLAYLPLDVKAPAARIDAIVSSLPGNKLILLGANVTPPKLQEAADIVFPIRD 4680
QY 130 ----- 129
Db 4681 TFFTLLDGLTQDGTIERPSAQSLAYAMFTSGSTGRPKGVWVQHNRIVRLVKNSNVAKQ 4740
QY 130 ----- 129
Db 4741 PAAARIAHISNLAFASSWEIYAPLNGGAIVCADIYFTTIDPQALQETFOBEIRGAMLP 4800
QY 130 ----- 129
Db 4801 PSLKQCLVQAPDMISRLDILFAAGDRPSSVDALQAOQLVSGGVFNAYGPTENTILSTIY 4860
QY 130 ----- 129
Db 4861 NVAENDSPVGPICSAVNSGAYIMDKNQQLVPAGVNGELVVTGDLARGYMDPKLDAD 4920
QY 130 ----- HTAKTITFRP ----- 139
Db 4921 RFIQLTVNGSEQVAYRTGDRVRYPKDFQIEFFGRMDQKIKIRGHRIEPAEVEQAFI 4980
QY 140 ----- 139
Db 4981 GFVEDVAIVIRTPENQPEMVAFTAKGDNARBEEATQIEGWEAHPGEGGAYANIEE 5040
QY 140 ----- PPCGR ----- 144
Db 5041 SEALGYDFMGWTSYDGTEDKDEMREWLNDTMRSLLDGKPAGRVLEVGTCGMIMFNLG 5100
QY 145 ----- 144
Db 5101 RSQGLERYIGLEPAPSAAEFVNNAKSPGLAGRAEVHVGTAADVGTQLGLTSDMAVINS 5160
QY 145 ----- VPSM ----- 148
Db 5161 VAQYFPTPEYLAETIKSLVQPGMKRIYLGDMRSWAMNDFAAARAYSLADNASKDRVR 5220
QY 149 ----- TCLSEML ----- 155
Db 5221 QKMELEKEEELLVDPAFTALASQLQDRIOHQVEIILPKMKATNELSSRYAAVLHISD 5280
QY 156 ----- NVSKRNDTGE ----- 165
Db 5281 EPLPIYKIDPEAWINPEGRSLTRELAQVLKENENAESVAISNIPYSKTVVVERHIVRSLD 5340
QY 166 ----- 165
Db 5341 QEDANAPESMGSDWISAVRTRAQQCHTILSASDLPIAEDAGPRVEVSUARQHSQHAL 5400
QY 166 ----- QG ----- 167
Db 5401 DAVPHLHKPATEDSRVLKPTDQGRPLKSLTNQPLLPQASRRRAELLIREGLQTLPPY 5460
QY 168 ----- CGNFT ----- 172
Db 5461 MIPSQITLIDRMPLNANGKVDRELARRAKITOKSKPVEDIVPPRNSVEATVCKGFTDVL 5520
QY 173 ----- TFPNMFNV ----- 181
Db 5521 GVEVGITDN--FFNLGHSMLATKLAARLGRQLNTRISVRDVPQPVVADLAAVIQNSA 5578
QY 182 ----- 181
Db 5579 PHEPIKPADYTGVPVQSFAGRLWFLDQNLVNGATWYLMPLGIRLHGLSLRVDALATAISAL 5638

QY 182 ----- 181
Db 5639 EQRHEPLRTTFHEBDGVGVQVVDHRRPKDLRIIDLSTOPKDAYLAVLKHEQTTLFLATE 5698
QY 182 PRWNTKL----- 188
Db 5699 PGWRVALIRLGEERHILSIUMHHIISDGWSVEVLFDENHRFYSSALROODPMEQILPLPI 5758
QY 189 ----- 188
Db 5759 QYRDPAAWQKTEQVABHQRLDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIEG 5818
QY 189 ----- 188
Db 5819 RLHDKLRAFRCVHOATPFVILLAAALRAAHRYLTGAEDATLGTPIANRNRPELENMIGPFV 5878
QY 189 ----- 188
Db 5879 NTQCMRTAIBENDNFESLVRVRSTATSAFANQDVPFESIVSSILPGSRDASRNPVQVI 5938
QY 189 ----- 188
Db 5939 LAVHSQODLGLKLTLEGLRDEAVDSAISRTRPDVEPHLFEHADRLSGSVLYAKELFKLRTIE 5998
QY 189 ----- 188
Db 5999 SVSVFLETLRALDQPLTPLAVLPLTDGVGEIASKGLLDVPRDYPRDANIVEFQOHV 6058
QY 189 ----- 188
Db 6059 RATPDATAVKDATSILTYAQLDQOSDLAIWLSRRHMPETLVGLVAPRSCETIAMPFI 6118
QY 189 ----- 188
Db 6119 MKANLAYPLDINSAPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOEILAGT 6178
QY 189 ----- YV----- 190
Db 6179 GLDKTGSNARPSATSLAYVIFTSGSTGPKGVNVEHRSVTRLAKPSNVISKLPGARVA 6238
QY 191 ----- 190
Db 6239 HLANIAFDASIEIATLLNGATLVCLDYHTVLD CRTLKEVFERESTVTVTLMALLKQC 6298
QY 191 ----- GPTKVNVDQSOTIY----- 203
Db 6299 VAEIPETLAHLDLLYTCGDRVGGHDAMRARSVKIGMFSGYGPENTVIS-TIYEVDAD 6357
QY 204 ----- 203
Db 6358 MFVNGVPIGTVNSGAYVMDRNQQLVPSGVVGVGLVVTGDLARGYTDPSLNKNRFIYT 6417
QY 204 ----- 203
Db 6418 VNGESTRAYTGRVRYRPHDLQIEPFGRMDQVKIRGHRIEPEGEVESALLSHNSVDAA 6477
QY 204 ----- 203
Db 6478 VVICAPADQDSGAEMVAFVAARNTEDEDTQEEAVDQVQGWETHFETAAYSEVKDIRQSE 6537
QY 204 ----- FLGLTAL----- 210
Db 6538 VGNDFMGWTSMYDGEIDKTDMEHLNDTMRMLDAREPGHVLBIGTGTGMFMENLAKCP 6597
QY 211 ----- 210
Db 6598 GLQYGVFEPSKSAQFVNDAQSPALKDGRSIVHVGTATDINKAGPIQPLRVVINSVA 6657
QY 211 ----- 210
Db 6658 QYFPTPEYLPVRVEALVQIPSVVERIVPGDMRTNAINRDFVASRALHTLGEKANRLVRQM 6717
QY 211 ----- 210

Db	6718	IYELEANE	BELLTDPAFFTSLRGLGEKIKHVEILLPKTMKATNELSKYRAAVLHVRGSR	6777	Db	7798	AETPELVAN	LEILHTAGDRLPDGDANLAGKTAKGGIFNVILGHTENTAYSTFYFVVBGETF	7857
Qy	211	-----	-----	210	Qy	240	ING-----	-----	242
Db	6778	EQSTIHQV	SPNAWIDFADAGLDROTLINLLKEHKDAGTVAIGNIPYSKTIVERFVNKSLS	6837	Db	7858	VNGVPGRG	ISGNHAYIIDRHQKLPAGVMGELITGDGVARGYTDTSALNKDRFYIDIN	7917
Qy	211	-----	-----	220	Qy	243	-----	-----	242
Db	6838	EDMEEGQ	NSLDGSAWVAVRMAQSCPSLDAMDVKEIAQEAQYQVEVSWARQWSQNGAL	6897	Db	7918	GKSTWSYR	TGDKARYRPRDGOLEFFGRMDQVKIRGVRIEPEGEVELTLLDHKSVAATVW	7977
Qy	221	-----	-----	220	Qy	243	-----	-----	242
Db	6898	DAIFHHF	PEPKEGARTLIEFPTDYBGRNVNTLTNRPLNSIQSRRLGTQIREKLQTLPPY	6957	Db	7978	VRRPENG	PEMIAFITIDAEDDVQTHKAIYKHLQILPAYMIPSHLVILDDQMPVDNGKV	8037
Qy	221	-----	-----	220	Qy	243	-----	-----	242
Db	6958	MIPSRIM	VLDQMPVNNNGKIDRKELVRRRAIVAPKPSRAATRVAPRNEIEAILRDEFEDVL	7017	Db	8038	DRKDLAL	RAQTVQKRSTAAARVPPRDEVEAVLCBEYSNLLVEVGITDGFDFLGHSLA	8097
Qy	221	-----	-----	222	Qy	243	TKLK-----	NT-----	248
Db	7018	GTEVSVL	DNFPLDGLGSHLMATKLAARVSRRLDAHISIKDVDPQVLAADLAASIQRESAPH	7077	Db	8098	TKLAARL	SRQLNTRVSKDVPDPILADLADIIRGSHRHDPIPATPYTGPVEQSFAQGR	8157
Qy	223	-----	-----	222	Qy	249	-----	-----	248
Db	7078	EPQPORY	TGPAEQSFAQGRWFLDQLNLGATWYLMPLAIRIGQLRVAALSAAALFALER	7137	Db	8158	LWFLEQL	NLGASWYLMPPFAIRMRGPLQTKALAVNALVHRHEALRTTFEDHDGVGVQVI	8217
Qy	223	-----	-----	222	Qy	249	-----	-----	257
Db	7138	RHETLRT	TFERSDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLRREQTVPFDLSEPG	7197	Db	8218	QPKSSQ	DLRIIDLSDAVDDTAYLAALKRBQTTAFDLTSEPCWRVSLRLGDDDDYILSVM	8277
Qy	223	-----	-----	222	Qy	258	-----	-----	262
Db	7198	WRVCLVK	TGEBDHVLSIVMHHIYDGSVDILRGELGQFYSAALRGQDPLHANPLIQY	7257	Db	8278	HHIISDG	WTVDVLREQELGQFYSAAIRGQEPISOAKSLPIQYRDPFVWQORQENIKEQAKQ	8337
Qy	223	-----	-----	222	Qy	263	-----	-----	262
Db	7258	RDFAWQ	REAKQVEHORQLGYMSKQLVDSTPAELLTDLRPSILSGRAGSDVDTIEGSV	7317	Db	8338	LKYSQOL	ADSTPCEFLTDLRPSILSGEADAVPMVIDGTVYQLLTDFCRTHQVTSFVL	8397
Qy	223	-----	-----	222	Qy	263	-----	-----	269
Db	7318	YGALQSF	CRTRSVTFVLLTVFRIAPHRLTAVDDATIGTPIANRNREPELTIVGCCFVNT	7377	Db	8398	LAAPRTH	YRLTGLDATVGTPIANRNREPELGLIGFFVNTQCMRMAISSETEFESLVQQ	8457
Qy	223	-----	-----	222	Qy	270	TQSTTTP	PFYS-----	279
Db	7378	QCMRISI	ADDONFEGVLQVRNVATAAVANQDVPPERIVSALVPGSRNTSRNPLVQLMFA	7437	Db	8458	VRUTTEA	FANQDVPPEQIVSTLLPGSRDTSRNPLVQVMEALQSQDQLRIQLEGMTDEA	8517
Qy	223	-----	-----	226	Qy	280	-----	-----	279
Db	7438	VQSVEDY	QVRLGLESVMPGEASTRDMEFHLVPGDQKLTGSVLYSSDLFEQGTIQNF	7497	Db	8518	LETPLST	RDLLEVHLFQEVGKLSGLLYSTDLFEVETIRGIVDVFLERLRRGLEQPKQL	8577
Qy	227	-----	-----	235	Qy	280	-----	-----	279
Db	7498	VDIFQEC	LSVLDQPLTPISVLPSNALSLESLLLEMPDSDYPRDRTVVDLFRQAAI	7557	Db	8578	MAMPITD	GITKLDDQGLLTIVAKPAYPRESSVIDLFRQQVAAAPDAIADWSSSTLTIVADL	8637
Qy	236	-----	-----	235	Qy	280	-----	-----	279
Db	7558	CPDSIAV	KDSSSOLTQAQLDEQDRVAALWHERHMPAESLVGVLSRPSRCTIIAYFGIMK	7617	Db	8638	DGQSNKL	AHLWCORNMAPETLVAVFAPRSCTTIVAFGLVKLANLAVILPDVNAPARIEA	8697
Qy	236	-----	-----	238	Qy	280	-----	-----	279
Db	7618	ANLAYPL	DVVYAPDARLAAILDTVEGERLILLGAGVPQPGIQIPRLSTAVIAEALSHATT	7677	Db	8698	ILSAVFG	HKLVLVQAHGPELGLTMDTELQIDEALASSSGDHEQIHASGPTATSLAYV	8757
Qy	239	-----	-----	238	Qy	280	-----	-----	279
Db	7678	VDVTSI	PQSATSLAYVIFTSGTGKPGVMIHRGIVRLVDRDNTVNVFPESGALPVSH	7737	Db	8758	MFTSGTG	KPGKGMIDHRSIIRLVKNSDVVATLPTPVPMANVSNLAFDISVQBIYTALLN	8817
Qy	239	-----	-----	238	Qy	280	-----	-----	279
Db	7738	FSNLAWD	AATWEIYTAVINGTGVVICIDRDTMLDIAALNSTFRKENVRRAAFPTFAFLKQCL	7797	Db	8818	GGTIVCL	DVLTLLDSKILYNVFVEAQVNAAMFTPVLLKQCLGNMPAIISRLSVLNVGDR	8877
Qy	239	-----	-----	239	Qy	280	-----	-----	279
					Db	8878	LDADAVA	ASGLIQDVAVYNAYGPTENGMOSTMWYKVDVNEFFVNGVPIGRSITNSGAYVMD	8937

Qy	280	-----YTTSA-----	284
Db	8938	GNQVSPGVMEI VVTGDLGARGYDTSALDEDFVHVHTIDGEEENIKAYRTGDRVYRPK	8997
Qy	285	-----	284
Db	8998	DFEIEFFGRMQQVKIRGHRIEPAVEHALLGDLVHDAVVLKPKANQBEMIAFITSQ	9057
Qy	285	-----	284
Db	9058	EDETIQESHNKQVQGWGEHFDVSRYADI KDLDTSTFGHDFLGWTSMDYDGDIPNEMKE	9117
Qy	285	-----	284
Db	9118	WLDETTASLDRNRPBGHILEIGAGTGMLSNLKGVDGLQKVGLDPAPSAIAIFVNEAVKS	9177
Qy	285	-----	284
Db	9178	LPSLAGKARVLVGTALDIGSLDKNEIOPELVVINSVAQYFTSEYLIKVKAVVEVPSVK	9237
Qy	285	-----ALN-----	287
Db	9238	RVFFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAELESEEEELLVDPAPFVSLSRS	9297
Qy	288	-----	287
Db	9298	QLPNIKHVEVLPKLMKATNELSSYRYAAVLHISHNEERQLLIQDIDPTAWVDFAATQKDS	9357
Qy	288	-----VTTNVTYSTTAARV-----	303
Db	9358	QGLRNLQOGRDDWMIAGNIPYKTIIVERHIMNSLDQDHVNSLDGTSWISDARSAAAIC	9417
Qy	304	-----	303
Db	9418	TSPDAPALTQAKBEGFRVELSWARQSRQNGALDAVFHRLATDANCERSRVLVHFPTDHO	9477
Qy	304	-----	303
Db	9478	GRQLTLNRPLQRAQSRRIEQVFEALQALPAYMIPSRILVLPQMPNTANGKVDKQL	9537
Qy	304	-----	303
Db	9538	ARRAQVAKRKAVSARVAPRNDTEIVLCEYADILGTEVGITDNPFDMGHSLMATKLAA	9597
Qy	304	-----	303
Db	9598	RLSRLDTRVTVKEVDFKPVLDLAASTEQGSTPHLPASSVYSGPVEQSYAQGLWFLD	9657
Qy	304	-----	303
Db	9658	QFNLNATWYHMSLAMRLIGPLNMDALDVALRALGQRHETLTTFEAQKDIGVQVVHEAGM	9717
Qy	304	-----	303
Db	9718	KRLKVLDSLDRKEHMAVLENEQMRPFTLASPCWKHGLARLGTEYILSLVMHMFSD	9777
Qy	304	-----	303
Db	9778	GWSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDPFAAWQKEAAQVAEHERQLAYWEN	9837
Qy	304	-----	303
Db	9838	QLADSTPCELLTDPRPQFLSGKAGVIVPTIEGFPVYEKLLKFSKERQVTLFSLVLLTAPRA	9897
Qy	304	-----STST 307-----	307
Db	9898	THPRLTGAEDATIGTPIANRRNPELEHIIGFFVNTQCMRLLLDTCSTFESLVQHRVSAT	9957
Qy	308	IAY-----	310
Db	9958	DAYSNDQIPFERIVSALLPGSRDASRSPLIQLMFALHSQDPLGNITILEGHEHRLPTSVA	10017
Qy	311	-----	310
Db	10018	TRFDMEPHLFOEPNKLGSILFADELFOPETINSVVTVQBILRRGLDQPOVSIETMPLT	10077
Qy	311	-----	310
Db	10078	DGLIDLEKLGILLBIESNFPDYSVDVPROQVAANPNAPAVVDSETSMSYTSLDQKSEQ	10137
Qy	311	-----RPDS-----	314
Db	10138	IAAMLHAQGLRPESLICVMAPRSPETIVSLFGLKAGYAVLPDVNSPAARIQIPLSEVE	10197
Qy	315	-----	314
Db	10198	GKRLVLLSGIDMPQSDRMDVETARIQDILTNTKVERSDMPRSATSLAYVIFTSGSTG	10257
Qy	315	-----	314
Db	10258	RPKGMIEHRNLRVLKQSNVTSOLPQDLRMAHISNLAFDASIWEIPTAILNGGALICID	10317
Qy	315	-----SPMK-----	318
Db	10318	YFTLLDSQLRRTTEKARVNATLPAPALLKECLNHAFTLPEDLVLYYIGDRLDATDAK	10377
Qy	319	-----	318
Db	10378	IQALVKGTVYNAYGFTENTVMSTIYRLTDGESYANGVPIGNAVSSSGAYIMDKORLVPP	10437
Qy	319	-----	318
Db	10438	GVMGELVSGDGLARGVTNSTLNADRFVDIVINDOKARAYTGDRTRYRPKDGSIEFFGR	10497
Qy	319	-----	318
Db	10498	MDQVKIRGRHVEPAEVEQAMLGKNAIHDAAVVQAVDQGETEMIGFVMSASDRFSEGE	10557
Qy	319	-----	318
Db	10558	EITNQVQEWEDHFESTAYAGIEAIDQATLGRDFTSWTMYNGNLIDKAEMBEWLLDWTQS	10617
Qy	319	-----	318
Db	10618	LLDKEDARPCAEIGTGTGMVLFNLPKNDGLESYGIEBRSRALLFVDKAAQDFPGLQKT	10677
Qy	319	-----	318
Db	10678	QILVGTAEIDIKLVKDHPDPVVVINSVAQYPPSRSYLVQIASELIHMTSVKTIFFGDMRSW	10737
Qy	319	-----	318
Db	10738	ATNRDFLVSRALYTLGDKATKQDQIRQEVARLEENEDELLVDPAPFTSLTSQWPGVKHVE	10797
Qy	319	-----	318
Db	10798	ILPKRMTSNELSSRYAAVLHICRDGEGRRNRYRRVHSVEENAWIDFASGMDRHALVQ	10857
Qy	319	-----	318
Db	10858	MLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQDSLDSAWQSAWKAMAARCP	10917
Qy	319	-----	318
Db	10918	CLSVTELVEIQAGFRVEVSUARQSRQHGALDVVPHLEDDRVGRVLNFTDFERLPP	10977
Qy	319	-----	318
Db	10978	STGLTSRPLORIQNRRFESQIREOLQTLPLPYMVPSRIVVLERMPLNANSKVDKRELARK	11037
Qy	319	-----SIMATOL-----	325
Db	11038	ARTLOTIKPSATRVAPRNDIEAVLCEFOAVLGVTVGVMDNPFELGGHSLMATKLAARLS	11097
Qy	326	-----	325

Db	11098	RELDRVSVKIDFNQPILODLADVQTGSAPEAIPTSPYSGPVEQSQRWFLDQLN	11157	Db	12178	VRDVEDTPVVGQAAASIOQGSTPHEAI	PALSHSGPVQSQFAGRLWFLDRNLNRAWTIM	12237
Qy	326	-----	325	Qy	357	-----	MKNTHVLIR	365
Db	11158	LNASWYHPLASRLRGLRIEALOSALATIEARHESLTTTPEQDGVVQIVRAARNKQL	11217	Db	12238	PFQVRLRGLRVDALQOTALRALERHELLRTT	FEQDQGVGMQIVHSPRMRDICVVDISGA	12297
Qy	326	-----	325	Qy	366	NE	-----	367
Db	11218	RIIDVSGTEDAYLAALKQEQDAAFDLTAEPCGRVALLRLGLDHRVLSIVMHHIISDGWSV	11277	Db	12298	NEDLAKKEEQAPPNLSTEVAWRVALFKAGENHHILSI	VNHHIISDGWSVDIFQOELAQ	12357
Qy	326	-----	331	Qy	368	-----	-----	367
Db	11278	DILRQELQLYSNASSOPAPLPIQYRDFAIWQKODSOIAEHQKQLNWKRLVNSKPAEL	11337	Db	12358	FYSVAVRGHDPLSQVKPLPIHYRDFAVWQRQDKQVAVHESOLQYWIQLADSTPAEILSD	12417	
Qy	332	-----	339	Qy	368	-----	-----	367
Db	11338	LADTPRKALSGDADVIPIEIDDOVQYNLRSFCRHRVTSFVALLAFAAHYRLTGAED	11397	Db	12418	FNRPEVLGSEAGTVPIVEDEVEYEKLSLFCRHNQVTSFVLLAARVAHVHRLTGAEDATI	12477	
Qy	340	-----	339	Qy	368	-----	-----	367
Db	11398	ATIGSPIANRRPELEGLIGFVNTQCLRIPVKSEDTFDFLVKQARETATEAQDNQDVPF	11457	Db	12478	GTPIANRRPELEDLIGFVNTQCMRIALEBHNFLSVRRVRSTAASAFENQDVPERL	12537	
Qy	340	-----	343	Qy	368	-----	TPYT	371
Db	11458	ERIVSSMVASSRDTSRNPLVQVMPFAVHSOHDLGNIRLEGVGPVSMAASTRFDEMHLP	11517	Db	12538	VSALLPGSRDASRNPLVOLMPVHVSQRNLKQLEGLEGEPTTYTATTRDFVPHLFEQD	12597	
Qy	344	-----	343	Qy	372	-----	-----	371
Db	11518	EDQMLGNVNVFSKDLFSEITIRSVAVFQETLRRGLANPHANLATPLTDGLPSRLSLC	11577	Db	12598	KGLAGNVVFAADLFEAATIRSVVEFHEILRRGLDQPDIAISTMPLVDGLAALNSRLFA	12657	
Qy	344	-----	343	Qy	372	-----	YGTLD	377
Db	11578	LQVNPDPDASVIDVFEQVASIPKSIADVIDASSQLTYTELDERSSQATLWLRQVTV	11637	Db	12658	VEDIEPDPATEASVVDVQTVVQVVPALAVTDTSTKLTAYELDQOQSHVAWLSKQKLP	12717	
Qy	344	-----	345	Qy	378	-----	-----	377
Db	11638	PEELVGLAPRSCETIIAFLGIKANLAYPLDVNAPAGRIETILSSILPGNRLILLGSDT	11697	Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMDNSVPEARQAILSEIPGEKPVLLGAGV	12777	
Qy	346	-----	345	Qy	378	-----	-----	377
Db	11698	QAVKLHANSVRFRISDALVESGPPTEBELSTRPTAQSLAYVMFTSGTGVPGVMVEHR	11757	Db	12778	PIPDNKTADVVMFISDIVASKTDSYSPGTRPSASSLAYVIFTSGTGRPKGMVMEHRG	12837	
Qy	346	-----	345	Db	12838	VISLVKQNASRIPQSLRMAHVSNLAFDASVWEIFTLLNGTLFCISYFTVLDLSKALSAA	12897	
Db	11758	GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGTGVVCIDYTTIDIKAL	11817	Qy	378	-----	MSSLY	384
Qy	346	-----	345	Db	12898	FSDHRINITLLPPALLKQCLADAPSVLSLESLEYIGDRLDGADATKVKDLVKKAYNAY	12957	
Db	11818	EAVFKOHHIRGAMLPALLKQCLVSAPTMISSLEILFAAGDRLSSQDAILARRAVGSGVY	11877	Qy	385	-----	-----	384
Qy	346	-----	345	Db	12958	GPTNSVMSTIYTIETHEFANGVPITGSLGPKSKAYIMDQOQLVPAGVMGELVAGDGL	13017	
Db	11878	NAYGPTENTVLSTIHNIGENAFSGVPIGNVANSAGFVMDQOQLVSAGIVELVVTG	11937	Qy	385	-----	E	385
Qy	346	-----	345	Db	13018	ARGYTDPSLNTGRFTHITIDGKQOAVRTGDRVYRPRDYQIEFFGRLDQKIRGHRIE	13077	
Db	11938	DGLARGYTDKSLVRDFIYITLDGNRVRAVRTGDRVRHRPKDQIEFFGRMDQKIRGH	11997	Qy	386	TMFVENKTASDS	-----	397
Qy	346	-----	345	Db	13078	PAEVEQALLSDSSINDAVVSAQNKEGLEVMVGYYITTQAAQSVDKKEASNKVQWEAHFDS	13137	
Db	11998	RIEPAVEQALARDPAISDAVITOLTDEBEPELVAFPSLKNANGTNGVNGVSDQEKID	12057	Qy	398	-----	-----	397
Qy	346	-----	353	Db	13138	TAYANIGGIDRDALQDQFSLSWTSMYDGLSIPREBQEWLNDTWESLLDNQPGKVLIGT	13197	
Db	12058	GDEQHALLMENKIRHNLQALLPTWIPSRIIHVDOLPNVANGKIDRNLAVRAQAPTPTS	12117	Qy	398	-----	NKTTTSPSM	407
Qy	354	-----	356	Db	13198	GTGMVFLNLGVEGLQSVAGLEPSRSVTAMWNKAIEFTFPLAGSARVHVGTAEIDISSIDG	13257	
Db	12118	SVSTYVAPRNDIETIICKFADILSVRGITDNFDFDGGHSLIATKLAARLSRLDTRVS	12177	Qy	408	-----	GFORTE	413
Qy	357	-----	356	Db	13258	LRSDLVWINSVAQYFPPSREYLAEITANLIRLPGVKIRIFFGDMRTYATNKDFLVARAVHTL	13317	

```
Qy 414 ----- 413
Db 13318 GSNASKAMVROQVAKLEDBEELLVDPAFTSLSDQFPDEIKHVEILLPKRMAATNELSSY 13377
Qy 414 ----- 413
Db 13378 RYAAVHVGHQMPNGEBEDKQWAVKOINPKAVDPFAGTRMDRQALLQLQDRGRDDVV 13437
Qy 414 ----- 413
Db 13438 AVSNIPYSKTIHERLSQSLDDSDGTSAVDGTAWISRTQRAKECPALSVDADLIEIGK 13497
Qy 414 ----- 413
Db 13498 IGFEVZASWQHSQSGLDVAFHRFBPPRHSGHVMFRFPTHEKGRSSSLNRPPLHLQ 13557
Qy 414 ----- 413
Db 13558 SRRLEAKVRERLQSLPPYMIPTSRITLLDQMLTNGKVKDRKRLARQARVIPRSAATLD 13617
Qy 414 ----- 413
Db 13618 FVAPRTIEVVLCSEFTDLLGVKIGITDNPPFELGHSLLATKLSARLSRRLDAGITVKQV 13677
Qy 414 ----- 418
Db 13678 FDQPVLAADLAASILQSSRRHSIPSLPYEGPVEQSFAGRLWFLDQFNIDALWYLIPFAL 13737
Qy 419 ----- 418
Db 13738 RMRGLOVDALAALVALEERHESLRTTFEERDGVGIVQVQPLRTTKDIRIIVSGMRDD 13797
Qy 419 ----- 418
Db 13798 DAYLEPQKEQOTPDPLASEPGWRVALLKLGKDDHILSIVMHHIISDGNSTEVLRBLGQ 13857
Qy 419 ----- 419
Db 13858 FYLAASKGAPLSQVAPLPIQYRDFAVWQREQVAESQRQDLYWKQLADSSPAELLAD 13917
Qy 420 Y----- 420
Db 13918 YTRPNVLSGEAGSVSFVINDSVYKSLVFCRSRQVTTFTLLAAFRAAHYRMTGSDDATI 13977
Qy 421 ----- 420
Db 13978 GTPIANRRPELENLIGGFVNTQCMRITIGDDETFESLVQQVRSTTATAFENQDVPFERI 14037
Qy 421 ----- 426
Db 14038 VSTLSAGSRDTSRNPVLQLLFAVHSQQGLGRIQLDGVVDEPVLSTVSTRPDEPHAFQEA 14097
Qy 427 ----- 430
Db 14098 DRLNGSVMFATDLFPETIQGFVAVVEVLQGLEQPSPIATMPLAEGIAQLRDAGALQ 14157
Qy 431 ----- 435
Db 14158 MPKSDYPRNASLVDVFOQQAMASPTVAVTSTSKLTYAELDRLSQQAAYLRRLQQLPAE 14217
Qy 436 ----- 437
Db 14218 TMVAVLAPRSCETIIAFLAILKANLAYMPLDNTFSPARMEAIISVPCRRLLLVGSGVRH 14277
Qy 438 ----- 437
Db 14278 ADINVPNAKTMVISDVTGTDAIGTPEPLVVRPSATSLAVVFTSGTGKPKGVMEHRA 14337
Qy 438 ----- 437
Db 14338 IMRLVKDSNVVTHMPPATRMARHVTNIAFDVSLFEMCATLNGTGLVCIDYLLDSTMLR 14397
```

```
Qy 438 ----- 437
Db 14398 ETFEREQVRAAIPPPALLRQCLVNMPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
Qy 438 ---PT----- 446
Db 14458 AYGETENAIISTIYNIDKHDPYVNGVPIGSVNSGAYVMDRNOQLLPFGVMBELVVTGE 14517
Qy 447 ----- 446
Db 14518 GVARGYTDASLDTRFVTVTIDGQRQAYRTGDRVVRPKGFQIEFFGLRQDQAKIRGHR 14577
Qy 447 ---EH----- 448
Db 14578 VELGEVEHALLSENSVTDAAVLRTMBEEDPQLVAFVTTDHEYRSGSNEEDPYATQAA 14637
Qy 449 ----- 459
Db 14638 GDMRKRLRLSLLPYVWPSRVITILRQMPNLNANGVKDRKDLARRAQMTPTASSSGPVHVAPR 14697
Qy 460 ----- 459
Db 14698 NETEAAICDFETILGVKVGITDNFPFELGHSLLATKLAARLSRRMGLRISVKOLFDDPV 14757
Qy 460 ----- 459
Db 14758 PVSILAGKLEQQQSGFGEDESSTVGIVPQLLPAEMSREIIQDVVPOIENGHSTPLDMYP 14817
Qy 460 ----- 459
Db 14818 ATQTIFFLHDKATGHATPPLFSLDFPETADCRRLASACALVQHFDIFRTVFSRGR 14877
Qy 460 ----- 459
Db 14878 FYQVVLHLADVPVEVETEQELDEVALHEADKQPLRLGRAMLRITAILKRPCKAKMLV 14937
Qy 460 ----- 461
Db 14938 LRMSHSLYDGLSLEHIVNALHALYSDKHLAQAPKFGLYMHMHASRRABGYNFRSILQGS 14997
Qy 462 ----- 462
Db 14998 SMTSLKRSVGALEAMTSPSAGTWTQTSKIRIPPAALKNGITQATLPTAAVSLLLAKHTKST 15057
Qy 463 ----- 462
Db 15058 DVVFGRVSGRQDLSINQDIVGPCINEVPVRVIDEGDDMGLLRATIQDQVTSFRHET 15117
Qy 463 -----W 463
Db 15118 LGLQEVKENCTDW 15130
```

RESULT 4

T30192

probable peptide synthetase - Aureobasidium pullulans

C:Species: Aureobasidium pullulans

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004

C:Accession: T30192

R:Peery, R.B.; Thornnewell, S.J.; Tobin, M.B.; Skatrud, P.L.

submitted to the EMBL Data Library, January 1997

A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureoba

A:Reference number: 220767

A:Accession: T30192

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-10797 <PEE>

A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:G4099310; PID:G4099313; PIDN:AA0005

C:Genetics:

A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2

C:Keywords: carrier protein

F:1618-1688/Domain: acyl carrier protein homology <ACPI>

F:3682-3752/Domain: acyl carrier protein homology <ACP2>

F;5615-5685/Domain: acyl carrier protein homology <ACP3>
F;7503-7573/Domain: acyl carrier protein homology <ACP4>
F;9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 22.6%; Score 555; DB 2; Length 10797;
Best Local Similarity 2.2%; Pred. No. 0.047;
Matches 227; Conservative 105; Mismatches 127; Indels 9990; Gaps 73;

QY	1	MGRKEM-----	6
Db	84	ISRKDLGIVPIYIPQCISLYISOLAILKSGAFCPLNLDVPGERLKFILKOTSASILLTT	143
QY	7	-----	6
Db	144	AAMRPKPLEGIVTVIVVDELSEFKSDREPAGEKITSPDTSSLAYIMYTSSTGLPKAV	203
QY	7	-----	6
Db	204	CLSHRAVTSQSLAHDRIPIFSRFLQFASPTFDVSVFEIFFPWPYRGATLVSVERNLLGD	263
QY	7	-----MVR---DVPKMFVLIS-----	19
Db	264	LPGTITSLNIDAELTPSVAASLVHRHENVPTLRALLTTIGEMLNTQVIQFGSGIKSGI	323
QY	20	-----	19
Db	324	LYGMVGPTEAIIHCTLPQGVGVDLPAGTIGIPLDTVSCFIVKPTSTKHASQLEILPICE	383
QY	20	-----	19
Db	384	IGELVIGHQLAGVYLNREEOTRAAFVTHPKFGLYRTGDKARLHRNGLTCEYGRISGQ	443
QY	20	-----ISFLVSPINC-----	30
Db	444	VKLGRQVRELGEIEHAASKAGGCHAVIASVISGLVLP--CIGDHRVSSKDIKSAACOKW	501
QY	31	-----KVMKAL---YN-----	39
Db	502	LPAYMIPSDIVLLDDFPYLPKGVKDKRLTSDYNSNTAQHVSGSDLSNAREITRIIES	561
QY	40	-----	39
Db	562	VLGVSDHSDTLSAAGLSLRAIQVASQLRRQGCADLGALELLSVSNVLADELVRKAD	621
QY	40	-----RP---41-----	41
Db	622	ESNINDNDSEKQVTHELRSSVERDPESKAPVSGIEDVLPTPLQDAMLVETAKRPOAY	681
QY	42	-----WRGLVLSKI--50-----	50
Db	682	CNELRTVSPKIPVERVRQALFALAQHRTALRSGFMPSGVSHCAVTVQVIMKTLVTSQFAH	741
QY	51	-----	50
Db	742	VKSFTTGSVNTRETLRLPLHFQYKCSGAEBILVAIHHALYDQMSVEVILEDLTLQN	801
QY	51	-----GKVKLD-----	56
Db	802	ERTPERPSGAVNKFFNLRRSEDQTSHLDFWGEVLSVTPGRLNLSFKMPPOPQSIQ	861
QY	57	---QLKLEILRQ-----	65
Db	862	HTIEMDETLRQAASHYSCSAHFVFAAYAILLGFYMGTEDTVFGTVFSGRTLPIVIES	921
QY	66	---LETTISTKYN-----	75
Db	922	MVGPLLSTLPIRINTLSRKPSDVLRLQEDNRKIMRHSMTSLADIKKAQFNPGEAFTD	981
QY	76	-----	75
Db	982	SIFVQETARPDAQAQTLLNLVEAHDYLEFNLTLELPTQQGVKTKATYQSSLLPLQHVK	1041
QY	76	-----VSKQ-PVKQNLMTNTEFPQYYIL-----	96

Db	1042	TLLQQLDALVKIVVARPETHMNEISDQLPISLVANSEPOSFVYKAGLSLVENHALNN	1101
QY	97	-----AGPIQNYGIT-----	106
Db	1102	SGGLALVFAHDIREGTSRMESLTYGELNTRANQLANLYLISOGAKRDELICVMEKSVSLY	1161
QY	107	-----	106
Db	1162	LSILAANKAGCGYLPVDPETPAARIQLAEADVFKCLTDSMAPVIADVSRCHIMNVD	1221
QY	107	-----YLMF-----	106
Db	1222	TDCSAQSGCTGQLPDFKPTDIAIYAVFTSGTTGPKGVLVTOENILSNLEVLSKIYVPEGS	1281
QY	107	-----	110
Db	1282	RLAQACNQAPDVSVFEIFFTWYTCMLCSASKDVMFRDPEKAINELITHLSLPTVAAL	1341
QY	111	-----	110
Db	1342	TDPAHIPRVKFLVTAGEAVTHHVHAGWAGKLYQGYGYPSETTNICTVNSAVESDHVINNI	1401
QY	111	-----	110
Db	1402	GPAFENTSAPVLTOGDDFQPLVPLGGLGELCFGGQQQVFRGYQNMPELTESKIINHNYGRI	1461
QY	111	-----	110
Db	1462	YRSGDLGRLPDGTILIQGRTDQKIRGQRIELGEISGCLQPPSPVQNCALIEVIKTADK	1521
QY	111	-----DFYS-----	114
Db	1522	ERLMAFWIPSGYKSDSYSILOPKNLEBIIKSIYAHADNLPAWMPDALVPVSAIPQTS	1581
QY	115	-----	114
Db	1582	QKIDKXERLASDGSALTVEDLNAYSRGADDDDETSLSATEQQLASALADTLOMSQTSIGR	1641
QY	115	-----TOLRKPACY-----	123
Db	1642	STSFALGLDSVAIRLATNLKREYGYSIDVSQILKRPTIARLAPLILGGSSKQTNBPVT	1701
QY	124	-----VYSOVNHTAKTIT-----	136
Db	1702	ADCEAAVGSYLHDESVVSQLHEHGQTVSQVLPTPLQEAMLSARDTSGSSAYRNKTLFSL	1761
QY	137	-----	136
Db	1762	HGSDVKLKACWEVLMQRHDLRTIFLSTEDSRFPFVQAVLSQWTLPWQECDDIPDQLSTL	1821
QY	137	-----	136
Db	1822	LDSAKAGDSIVDHPSPWKIQVYRSESTVYLLDMHHALYDANAMSNLLYEVEQLYKQDS	1881
QY	137	-----PRPPP-----	141
Db	1882	LSAPVSFKPLPNPMISTSVBEADALFRDQLRDEFVPKPRKRTDVKSFGFTITGRNLNYSKM	1941
QY	142	-----CGRV-----	145
Db	1942	VETFLSKHSTTMLSLTQAMMMKTLAASQSVSDVCCGVNVSGRSVPVDGIESLVAPCFNTI	2001
QY	146	-----	145
Db	2002	PVRVLSKXHSNGLVKALQVRNIDSLPYQLTPLRRIQAQAGTNGKRLFDLSVLQQDPT	2061
QY	146	-----	145
Db	2062	DLDSAIRLEGESGVMDDELILNPLTLLEGPPELLHELISGPETSCTALEYLDADOK	2121
QY	146	-----	145

Db 2122 TITSLYDLHRRSEHLAWRLIEVSRHSKODLGIPIYIPOCISLYISQALILKSGAFC 2181
QY 146 ----- 145
Db 2182 PLNLDPGERLKFILKDTASILLTTAAMRPKLPELEGITVIVDDELSEFKSDREPAGE 2241
QY 146 -----PSMTCLS----- 152
Db 2242 KITSPDTSSLAYIMYTSSTGLPKAVCLSHRAVTOSSLAHDRFIPFSRFLQFASPTPDV 2301
QY 153 ----- 152
Db 2302 SVFEIPFPYRGATLVSVERNRLGLDPLCTITSLNIDAEALTPSVAASLVRHRENVPTLR 2361
QY 153 -----EMLN----- 156
Db 2362 ALLTIGEMLNTQVIOQFGSGIKSGILYMGYPTEAAIHCTLPQPGVDLPAGTIGIPLD 2421
QY 157 ----- 156
Db 2422 TVSCFIVKPTSTKASQLEILPIGEIGELVIGGHQADGYLNREBQTRAAAFVTHPKFEG 2481
QY 157 ----- 156
Db 2482 LYRTGDKARLHNGTLECYGRISGQVKLRGORVELGEIEHAASKAGSCHAVIASVIGL 2541
QY 157 ----- 156
Db 2542 LVLFICGDPHRVSSKDIKSAQKMLPAYMIPSDIVLLDDFPVLPBGKVDKRLTDYNSN 2601
QY 157 -----VSKENDTGE-----QCGG- 169
Db 2602 TAQHVSGSSDISENAREITRIESVLGVSIDHSTDLSAAGLDSLRAIQVASQLRRQGCAD 2661
QY 170 ----- 169
Db 2662 LGALELLSVSNVALDELVRKADESINNDSEKWKQTVHELRSSEVERDFESKAFVSGI 2721
QY 170 ----- 169
Db 2722 EDVLPCTPLQDAMLVETAKRPOAYCNEURLTVSPKI PVERVROALFALAQHRTALRSFEM 2781
QY 170 -----NFTT-----FNM- 177
Db 2782 PSGVSHCAYQVIMKTLVTSQFAHVKSPTTGMVSNRETLLRPLHFQYKCSGABAEILVA 2841
QY 178 -----FENVPR-----WNTKL- 188
Db 2842 IHHALYDQWSEVEILEDLETLQNERTPERPSPGAVNKFFNLRSSEBQTSHLDFWGEYLS 2901
QY 189 ----- 189
Db 2902 DVTGRLPNLSPKMPPOPLQSIQHTIEMDMETLROAAHSYSCSAHVFFQAAVAILLGFY 2961
QY 190 VG-----PTKNV----- 197
Db 2962 MGTEDTVFGTIVSGRTLPIVEIESMVGPLSTLPLRINTLESRKFSVDLSRLQEDNRKIM 3021
QY 198 -----DS-----QT----- 201
Db 3022 RHSMTSLADIKKACGFNPGFAVFDISI FVMQETARPDARAQTLNLNVEAHDYLEFNLTLEL 3081
QY 202 ----- 201
Db 3082 EPTQGVKTKATQSSLLPLQHVKTLLQQLDALVKIVVAREPETHNNEISDQLPISVLSVA 3141
QY 202 -----IYFLGLTALLRYAORN----- 218
Db 3142 NSEPOSFVYKAGLGLSVENHALNNSGLLVFAHDI REGTSMESLTVGELNTRANQLAN 3201
QY 219 ----- 218
Db 3202 YLISQGAKRDELI CVMKEKSVLSYLISLA AVKAGCYLPLVPETPAARIQILAEADVKF 3261

QY 219 ----- 218
Db 3262 CLTDSMAPVIADVSRCHIMNVDTTDCSAQCTGPDLDKFTDIAYAVFTSOTGPKXGV 3321
QY 219 ----- 218
Db 3322 LVTOENILSNLEVLISKIYPVPEGSRLLOACNQAFDVSVEIFEPTTWTGMLCASKASQDWMF 3381
QY 219 ----- 218
Db 3382 RDEKAINELLEITHLSLUTPTVAULTDPAH IPRVKFLVTAGEAVTHHVHAGWAGKLYQGY 3441
QY 219 ----- 218
Db 3442 GPSETTNICTVNSAVESDHVINNIGAPAFENTSAFVLTOGDDFQLVPLGGLGELCFGGQV 3501
QY 219 ----- 218
Db 3502 FRGYONMPELTESKI INHPNYGRIYRSGDLGRLLPDGTILIOGRTDDQORKIRGRIELGE 3561
QY 219 ----- 218
Db 3562 ISGCLLOPPSVONCAIEVIKTADKERLMAFWIPSGYSKDSYSLIQPDKNLEILKSIYAH 3621
QY 219 ----- 218
Db 3622 LADNLPAYMVPDALVPVSAIPQTSQKIDKRRLASDGSALTVEDLNAYSGADDDDETSEL 3681
QY 219 ----- 218
Db 3682 SATEQQLASALADTLQMSQTSIGRSTSFALGLDSVAIRLATNLNRKEYGYSIDVSQILK 3741
QY 219 -----CT-- 220
Db 3742 RPTIARLAPLLGESSKOTNEPVTADCEAAVGSYLHDESVVSQLHEHCQTVSQVLPCTPL 3801
QY 221 ----- 220
Db 3802 QEAMLSARDTSGSAYRNKTLFSLHGSVDKLKACMEVMLQRHDILRTIFLSTEDSRPFV 3861
QY 221 ----- 220
Db 3862 QAVLSQNTLPMQECDDIPDQLSTLLDSAKAGDSIVDHPWPKIQVYRSESTVYLLDMH 3921
QY 221 HSPYLVNAMSRLPRV----- 236
Db 3922 HALYDANMSNLLYEVEQLYKQSLAPVSFKPFLNFMISTSVEEADALFRDQLREFVPK 3981
QY 237 -----PKYI-----CT-- 240
Db 3982 PKRTDVKSGFGTITGRNLNYSKPMVETFLSKHSTTMLSLTOAMMKTLAASQSYSDVCCG 4041
QY 241 ----- 240
Db 4042 NVWGSRPVVDGIESLVAPCFNTIPVRVDLSKHSRNLGLVKALQVRNIDSLPYQLTPLRR 4101
QY 241 ----- 240
Db 4102 IQAQAGTNGKRLFDLSVLLOQDPTDLSAIWRLEGESGVMDMRPKLPELEGITVIVUDE 4161
QY 241 ----- 240
Db 4162 LSRFKSDREPAGEKITSPDTSSLAYIMYTSSTGLPKAVCLSHRAVTOSSLAHDRFIPSF 4221
QY 241 ----- 240
Db 4222 SRFLQASPTFDSVFEIPFPYRGATLVSVERNRLGLDPLPGTITSLNIDAEALTPSVA 4281
QY 241 ----- 240
Db 4282 SLVRHRENVPTLRALLTIGEMLNTQVIOQFGSGIKSGILYMGYPTEAAIHCTLPQGF 4341

QY 241 ----- 240
Db 4342 VDLPGATTIGIPLDVTSCFIVKPTSTKHASQLEILPIGEIBELVIGGHQADGYLNREEQ 4401
QY 241 -----NGT----- 243
Db 4402 TRAAFPVTHPKFGLYRTGDKARLHRNGTLECYGRISGGQVKGQRVBLGIEHAASKAG 4461
QY 244 ----- 243
Db 4462 GCHAVIASVISGLLVLCIGDHRVSSKDIKSACQKWLPAWIPSDIVLLDDFPYLPSCG 4521
QY 244 ----- 243
Db 4522 VDKKRLTDYNSNTAQHVGSSDLSENAREITRIIESVLGVSIDHSTDLSAAGLDSLRAI 4581
QY 244 -----KLTMTMKLKKKQ 256
Db 4582 QVASQLRRQGCADLGALELLSVSNVLADELVRKAKADESNINDNSEKWKQTVHELRL 4638
QY 257 APVKEQFEKK-----AKTQS-----TTTP----- 276
Db 4639 SSVRFESKAFVSGIEDVLPCTPLQDAMLVETAKRPOAYCNEURLTVSPKIPVERVROA 4698
QY 277 ----- 276
Db 4699 LFALAQRHTALRSGFMPGSHCAVTVQVWKTVTSQFAHVKSFTTGWSVTNRRETLRLPL 4758
QY 277 YFSYTTSA-----LNV----- 285
Db 4759 HFQYKCSGAEABILVAIHALLYDQMSVEVILELTLQNBERTPERPSGAVNKFNLRR 4818
QY 286 ----- 285
Db 4819 SEDQTSHLDFWGEYLSVDTVPGRLPNLSPKMPPQPLQSIQHTIEMDMETLROAAHSYCS 4878
QY 286 ----- 285
Db 4879 AHVFPQAAVAILGPFMGCTEDTVFGTVFSGRTPIVEIESMVGPLLSTLPTINTLESRK 4938
QY 286 ----- 285
Db 4939 FSDVLSRLQEDNRKIMRHSMTSLADIKKACGFNGEAVFDSIFVMQETARPBARAQTLN 4998
QY 286 -----LNV----- 285
Db 4999 LVEAHDYLEFNLTELEPTQGVKTKATYQSSLLPLQHVKTLLQQLDALVKIIVARPETH 5058
QY 286 ----- 285
Db 5059 MNEISDQLPISVLSVANSEPOSFVYKAGLSLVENHALNNSGGLALVPAHDIREGTSRME 5118
QY 286 ----- 285
Db 5119 SLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLSILAANKVAGCGYLPVPET 5178
QY 286 -----LNV----- 288
Db 5179 PAARIQILAEADVFKCLTDSMAPVIAVDSRCHIMNVDTTDCSAQSCCTGFPQDFKPTDI 5238
QY 289 ----- 288
Db 5239 AYAVFTSGTGTKPGVLVTQENILSNLEVLISKIYPVPEGSRLQACNQAFDVSVEIFFT 5298
QY 289 ----- 288
Db 5299 WYTGMLCSASKVMFRDFEKAINELITHLSLPTVAALTDPAHPRVKFLVAGEAVT 5358
QY 289 -----TTNV----- 292
Db 5359 HHVGAWAGKGLYQGYGSEPTINICTVNSAVESDHVINNIGPAPENTSFAVLTOGDDFQL 5418
QY 293 ----- 292

Db 5419 VPLGGLGELCGGOQVFRGYQNMPELTESKIINHNPYGRIVRSGLDGLRLPDGTILIOQR 5478
QY 293 -----TYSI- 296
Db 5479 TDDORKIRGORIELGEISGCLLQFPFVQONCAIEVIKTADKERLMAFWIPSGYKDSYSIL 5538
QY 297 ----- 296
Db 5539 QPDKNLBEI KSIYAHILADNLPAWVPDALVPVSAIPQTSQKIDKRRLASDGSALTVED 5598
QY 297 -----TTAARRVSTS 306
Db 5599 LNAYSRGADDDTSELSATEQOLASALADTLQMSQTSIGRSTSPFALGLDSVAIRLATN 5658
QY 307 -----TIA----- 309
Db 5659 LRKEYGYSIDVSOILKRPTIARLAPLILGGESSKOTNEPVTADCEAAVGSYHLHDSVWSQL 5718
QY 310 ----- 309
Db 5719 HEHGQTVSQVLPCTPLQENMLSARDTSGSSAYRNKTLPSLHGSVDKLCACHEVMLQRHDI 5778
QY 310 -----YRPDSS----- 315
Db 5839 QVYRSESTVYLLDMHHALYDANAMNLLYEVOQYKQDLSAPVSPKPFNFIMISTSV 5898
QY 316 -----F 316
Db 5899 EADALPRDQLREFVPKPKRTDVKSFGTITGRNLNYSFKQVETPLSKHSTMLSLTQAMW 5958
QY 317 MKSIMATQ----- 324
Db 5959 MKTLAASQSYSDVCCGNWVSGRSVPVDGIESLVAPCFNTIPVRVDLSKHSRNLGLVKALQ 6018
QY 325 -----LR-----DLATW-----VYT 334
Db 6019 RVNIDSLPYQLTPLRRIOAQAGTNGKRLFDLSVLLQQDTTDLDSAIWRLEGESGVMDMYT 6078
QY 335 T----- 335
Db 6079 SGSTGLPKAVCLSHRAVTQSLAHDRFIPSPSRFLQFASPTFDVSVFEIPFPWYRGATLV 6138
QY 336 -----LRYRQNP----- 342
Db 6139 SVERNRLIGDLPGTITSLINIDAAELTPSVAASLVHRHENVPTLRALLTIGEMLTQVIOQ 6198
QY 343 ----- 342
Db 6199 FGSGIGSGILYGMGYPTAAIHCTLOPGFVDLPAGTIGIPLDTVSCFIVKPTSTKHA 6258
QY 343 ----- 342
Db 6259 SOLEILPIGEIBELVIGGHQADGYLNREBQTRAAFTVTHPKFGLYRTGDKARLHRNGTL 6318
QY 343 -----FC----- 344
Db 6319 ECYGRISGGQVKGQRVBLGIEHAASKAGGCHAVIASVISGLLVLCFICGDPHRVSSKD 6378
QY 345 ----- 344
Db 6379 IKSACQKWLPAWIPSDIVLLDDFPYLPSPGVKDKRLETDYNSNTAQHVGSSDLSENAR 6438
QY 345 ----- 344
Db 6439 EITRIIESVLGVSIDHSTDLSAAGLDSLRARIVASQLRROGCADLGALELLSVSNVLALD 6498
QY 345 ----- 344

Db 6499 ELVRAKADESININDSEKWKOTVHELRSSEVERDFESKAFVSGIEDVLPCTPLQDAMLVE 6558
Qy 345 ----- 344
Db 6559 TAKRPOAYCNELRLTVPKIPVERVRQALFALAQRHTALRSGFMPSGVSHCAYTQVIWKT 6618
Qy 345 ----- 344
Db 6619 LVTSQFAHVKSFTTGWSVTNRRETLRLPHQYKCSGABAILVAIHHALYDQMSVEVILE 6678
Qy 345 ----- 344
Db 6679 DLETLQNERTPERSFGAVNKFPNLRSEDOTSHLDFWGEVLSVDTFGRLPNLSPKMP 6738
Qy 345 ----- 344
Db 6739 POPLQSIQHTIEMDMETLQAAHSYCSARVFFQAAIILGFGYMGTEDTVFGTVFSQRT 6798
Qy 345 ----- 344
Db 6799 LPIVEIESMVGLLSTLPTPRINTLESRKFSVLSRLQBDNRKIMKHSMTSLADIKKACGF 6858
Qy 345 ----- 348
Db 6859 NPGEAVPDSIFVQETARPDAQAOTLLNLVEAHDYLFNLTLELEPTQOGVKTKATYQSS 6918
Qy 349 ----- 348
Db 6919 LLPLQHVKTLLQOLDALVKIVVARPETHMNEISDQLPISVLSVANSEBPQSFYKAGLSL 6978
Qy 349 ----- 348
Db 6979 VENHALNNSGGLALVFAHDIREGTSRMSLTYGELNTRANQLANYLISQAKRDELICVC 7038
Qy 349 ----- 348
Db 7039 MEKSVLSYLSILAAVACGGYLPVPETPAARIQILAEADVKFCCLTSSMAPVIADVSR 7098
Qy 349 ----- 348
Db 7099 CHINNVDTTDCSAQSGCTGQDPKPTDIAYAVFTSGTGKPKGVLVTOENILSNLEVLK 7158
Qy 349 ----- 348
Db 7159 IYPVEGSRLLQACNQAPDVSVEIFFTWYTMCLCSAKDVMFRDPEKAINLEITHLS 7218
Qy 349 ----- 358
Db 7219 LPTVAALTDPAH1PRVKPLVTAGEAVTHVHGAWAGKLYOGYGPSETTICTVNSAVE 7278
Qy 359 NTHVL----- 363
Db 7279 SDHVINNIGPAPENTSAPVLQGGDDFQLVPLGGLGELCFGGQVPRGYQNMPELTESKII 7338
Qy 364 ----- 363
Db 7339 NHPNYGRIYRSGDLRLLPDGTILTIQGRTDQDKIRGORIELGELGCLLOFPSPVONCAI 7398
Qy 364 ----- 363
Db 7399 EVIKTADKERLMAFWIPSGYSKDSYSILOPKNLEIITKSYAHADNLPAYMVPDALVP 7458
Qy 364 ----- 363
Db 7459 VSAIPQTSQKIDKRRLASDGSALTVEDLNAYSRGADDDTSELSATEQQLASALADTLQ 7518
Qy 364 ----- 363
Db 7519 MSQTSIGRSTGFFALGLDSVAIRLATNLKREYGVSDVSQILKRPRTIARLAPLLGESS 7578
Qy 364 ----- 364
Db 7579 KOTNEPVTADCEAAVGLHDESVVSQLHERGQTVSQLVPLCTPLQEAMLSARDTSGSSAY 7638

Qy 365 RNETPYTYITGTD----- 377
Db 7639 RNKTLFSLHSGVDLKLACWEVMLQRHDLRTIFLSTEDSRFPFVQAVLSQWTLPMQECDD 7698
Qy 378 ----- 381
Db 7699 IPDQLSTLLDSAKAGDSIVDHSPPWKIQVYRSESTVYLLDMHHALYDANAMSNLLYEV 7758
Qy 382 ----- 381
Db 7759 EQLYKQSLAPVSPKPLNFMISTSVEEADALFRDQLREFVPKPKRTDVKSFGTITG 7818
Qy 382 ----- 381
Db 7819 RLNYSPKMVETFLSKHSTTMLSITQAMMMKTLAASQSYSDVCCGCVSGRSVPVDGIESL 7878
Qy 382 ----- 381
Db 7879 VAPCNTIPVRVDSLKHSRNLGLVKALQRVNIDSLPYQLTPLRRIQAQAGTNGKRLFDL 7938
Qy 382 ----- 381
Db 7939 VLLQDQTTDLDSAIRWLEGESVMDPCIVELAPTNESYTLSLHFNRSYLDDEVVSNLHQ 7998
Qy 382 ----- 381
Db 7999 ACLSAFASCIRYPSDSVDFIDFADLVAGVLKPTDKHQPVEAAKTNRSEKSGSGDES 8058
Qy 382 ----- 381
Db 8059 WSPLELQIRAAYSAVSSAPEDRIRDDTTYKGLDLSIAIQANRLRKDGLLVQASDYM 8118
Qy 382 ----- 381
Db 8119 SPSCSELASAVQSRSTQVPLDERGDFEGFDKHYRGAALQSHRIATEKVASVRPCTPLQS 8178
Qy 382 ----- 388
Db 8179 GMLSEYTHSDGHQYFNHTFYAIEADIDSLSKLSAWSKVLQEHHELLRTGTVGTDDEHPFV 8238
Qy 389 ----- 388
Db 8239 MLTYTFDVIDLEIQASSREGSVYSEKASESVKNNLHLPMPWRSLLVGEGRCLOFS 8298
Qy 389 ----- 388
Db 8299 AHHAIPDAESLRLIMTDLQSALSNGYVPTRLTIDGALGHILSNQADVESQRTFWSQKLS 8358
Qy 389 ----- 388
Db 8359 GAPVTRPNNMTPVRISDTEAANVELVNLKRSKLEARCOELGVSMQSVGQAARLLSAY 8418
Qy 389 ----- 396
Db 8419 TGESQVTFGVVLSGRTSPATADAPPCITTLVPSTNTAVDDSQFLKOLMSYNATIQQHOF 8478
Qy 397 ----- 396
Db 8479 TPLTNIRYABSTSEALFDSLFTVYQRPNNVVDSSWKIIREKASVELASVEMEALSED 8538
Qy 397 ----- 396
Db 8539 GLGLRLTIDPAQVPHEQKIMLQQMEVMIAGLLKPEDAINTSVMSIIPPKPIIATDFKY 8598
Qy 397 ----- 401
Db 8599 LHTEASVKSYSRIAMEFYDALEDGQISRRHWTYRQDDEANKIAHLLIDRGVKGDI 8658
Qy 402 ----- 406
Db 8659 IATSFDKCPASPAFYGILKAGCAFCADPTAPARKAFILEDNARVLLTSDSIRSEL 8718

QY 407 ----- 406
Db 8719 ELTQCIIIDLINFKNELSTSSVPVGLAPSSVSVLYTSGTGTGPKGCEITHDNAVL 8778
QY 407 -MGQRTF----- 413
Db 8779 VMSFKRLFKGRWTDSEWLOFASVYHFDVSVLEQFWTWVGMRLVCAPRDLILEDIAGFLD 8838
QY 414 -----TDPL----- 417
Db 8839 TMQITHLDLTPSLGRLLDPALVPSLHKGVFITGESLKQDQINTWGDVGLCFNFYGFTEC 8898
QY 418 -----WD----- 419
Db 8899 TIGTVFPVCPKEGKPSNICQPDNVGCVYLAFTQTPVLRGALGELICISGLVKGKYL 8958
QY 420 -----YLSL-----LFLDEIR--- 431
Db 8959 RPELTADCFPYLDAGFGRVYRTGDLVRLFDHGSIDFLGRKNQVQLRGQRLEIDEIEAVI 9018
QY 432 ----- 431
Db 9019 KRCQDIQDTCIVAKHPKQDKQOLIAFIGINESRKQKPELCPAESTRHLIQTARAACE 9078
QY 432 ----- 431
Db 9079 RLFYGMVPTHFLPIQRIPLSVNNKVEKLLRLQYADLPTTVIQTATQADSQSLSDCEQ 9138
QY 432 -----NFS-----LRSPTVV 441
Db 9139 KVAQALABELLKIDDNDLTSSNIFSLGSSISAIOFSKKLKAGFTTVQVATVLKNPTIS 9198
QY 442 NLT----- 444
Db 9199 RLTKALATSGRSGGIBADAKQVISACRQRMGTVTVRLRCCKADIEAIPCTPQQGII 9258
QY 445 ----- 444
Db 9259 SRSLASESLYFNSFKFNAGVDLQKLGAFNAQLERTQILRTFETDDGVQAVRKTG 9318
QY 445 ----- 444
Db 9319 HLPWWTLEVYDLASVDGVPAKRQKWSYNTSHLTVPEIVIVRSGETFVSVDLHLY 9378
QY 445 -----pp 446
Db 9379 DGNFDFILMNVSKLYNSQEADFGKPFVDCLAFGLPLNVQAKFWLHLPDVKSASMP 9438
QY 447 ----- 446
Db 9439 LIDNPASHDVLCTASLDILNQADELRSLGVTVQALVQATWATLKRHYQGAIGTVWSGR 9498
QY 447 ----- 446
Db 9499 SIDFDGVENVIGLENTIPFYLRCEPGDTWTLVQRCHDFNTALPYQHTPLRDIVKCN 9558
QY 447 ----- 446
Db 9559 KHSRPLFDALFVYQGTLDNSDNTNHSILKPLEDDSFADYPLSPAEAEAGGNLKISVAA 9618
QY 447 ----- 446
Db 9619 KASICNETKARELIDEFHQAFAMNKSPENEVGASIGHFTPERLARTEDGTKEIQARDTS 9678
QY 447 ----- 446
Db 9679 DFHWGSEASVIRSEIATLAGVQEQADIDERTSIFEVLGDSVDVAVKLSRLKKGVLPVST 9738
QY 447 ----- 446
Db 9739 IMRSQTIAQMVLTLNGKTDTSKAKSGAFEALESKLTSTVARDQINGDKSFERVLPSPM 9798
QY 447 ----- 446

Db 9799 QBALVSEMVRTSYKAYFNHVDVLKLSKELDIVRLEQAWRTVIKVSPILRTGLEIDDDPDN 9858
QY 447 -----EHRRAVNLSTN----- 458
Db 9859 ATFAQIIQKAGFADPDHKLKSSDAINEHLAHINSVSSAALFQPLRLTRVSTPEDNYL 9918
QY 459 ----- 458
Db 9919 VLSVAHALYDGHSLSLHLDVQNAVNETFQSRPFVADILRDTFHGSNEESVMFWRSLTD 9978
QY 459 ----- 458
Db 9979 AKKTQLPOIKNEGSVTTTHRTKTSRLAANDLSSFCQKQGVTLQAACQATAWAFVLAHTV 10038
QY 459 ----- 458
Db 10039 QSLDVMFGVVLAGRDSLELAQVMPFMMNTVIMRSVLHGRKDVMSQIQATISDNHQHF 10098
QY 459 ----- 458
Db 10099 PLRQIQAACQGVQSTGQNDAPFDSLETFQRRPDTTSTKQMLYESVNGASEVEYPPVAVE 10158
QY 459 ----- 458
Db 10159 AEIVDGSILIWRAGKSSAFDEBEGVRNILDLTLEDVLQAIVTQPHEPTLAFDNQEVSI CDLP 10218
QY 459 ----- 458
Db 10219 AFQOQEKTAGSQTAGVSDSEI DEEHWSDTESAVREALAKVTKTPVQEISKTIASIQNLGV 10278
QY 459 -----SLW----- 461
Db 10279 DSINAIKISALLRQKIIIRVTVSEIVRAGSVWKIASVAQNKKOKKGHQAADKVIIDL 10338
QY 462 ----- 461
Db 10339 MAQKFTPDKFGYEDHQIEQILPATAGVYMLGAWQTNGQLFFGQFEYVTAVTMTDDI 10398
QY 462 ----- 461
Db 10399 KOAQKVTASNAVLRTVFVATODNETFVLQVLRLDAPTSFVDLDBETPASIIIGQFPVRL 10458
QY 462 ----- 461
Db 10459 TAQKVDDTYKLCIKIHHALYDAISLSLQELGNLYLQDSATTQOSPVKFADFAQPLGTV 10518
QY 462 -----WWL 464
Db 10519 DKTAKPFW 10527

RESULT 5

T31307
type I fatty acid synthase homolog - Cryptosporidium parvum
C;Species: Cryptosporidium parvum
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31307
R;Zhu, G.; Marchewka, M.J.; Woods, K.M.; Upton, S.J.; Keithly, J.S.
submitted to the EMBL Data Library, August 1998
A;Description: Characterization of a type I FAS gene in the parasitic protozoan Cryptosp.
A;Reference number: Z20993
A;Accession: T31307
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-8243 <ZHU>
A;Cross-references: UNIPROT:O96554; EMBL:AF082993; NID:g4092068; PID:g4092069; PIDN:AAC9;
C;Genetics:
C;Note: FAS1
C;Keywords: carrier protein
F;100-714/Domain: acetate-CoA ligase homology <ACL>
F;791-861/Domain: acyl carrier protein homology <ACP1>
F;906-1308/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:2924-2992/Domain: acyl carrier protein homology <ACP2>
 F:3062-3463/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:5157-5227/Domain: acyl carrier protein homology <ACP3>
 F:5274-5689/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>

Query Match 21.2%; Score 520; DB 2; Length 8243;
 Best Local Similarity 2.7%; Pred. No. 0.17; Mismatches 137; Indels 7589; Gaps 70;
 Matches 217; Conservative 100;

QY	3	RKEMVRDVPKM	-----FVLISISFL-----	23
DB	197	RKEEYINEVQKAINDIILNDSNHDFLGFDFNINVAFLQFTSGTSQPKGVMTGHNLL	: : :	256
QY	24		-----	23
DB	257	YNIHICISSYSPVLEBHYKNELDISIPAGITVLAKNELTEDELSFIHNEFCEKNIMSL	: : :	316
QY	24		-----LVSRI-----	28
DB	317	NDIKYDKLYLMSIINKVLYKNTMSSASVFWLFPVYHDMGLIGFCTPLPFGCNIQMS	: : :	376
QY	29		-----NCKV-----	32
DB	377	IDFIKPYLWQCMDKYNCSVGAENFAFEVVRKTPKSEILNQLNKLKIVFALLSGAEP	: : :	436
QY	33		-----	32
DB	437	KTIDRFEAFKSVGIKHNVIKPAYGLAEHTLIVAGSNSFOQEVNHTVNTKLRKNIV	: : :	496
QY	33		-----	32
DB	497	BIKEAKSSTETTSFVSGGIVYKIDIRIVNPESLKEVTPGNVGEIWISSSVTLGYN	: : :	556
QY	33		-----MSKALYNR-----	40
DB	557	NKEETKTFNANFTMLDCKTSKTYWRTGDSGFINDMLYISGRKDMIIIRGNFYQN	: : :	616
QY	41		-----	40
DB	617	IEEVIDGSGVRQGSVAVSVTQADGEAIGVAVEIRMETSIILGRVRRFFKEPAYDNIV	: : :	676
QY	41		-----PW-----	42
DB	677	AISKAVFTGHLGVHYIWLSPRTILKTSKIRRSQTRDAIFSKVLPLFEWCEQSPM	: : :	736
QY	43		-----RGLVLSKIG-----	61
DB	737	SDILVERSSQFISPTSMIISKGVQNKFKYEKTHSGQOKISBQANDEKNEFDIDQVK	: : :	796
QY	62	ILR	-----	64
DB	797	IEEPAQVLGITQKDLNAPLHEYGIDSLGSIIRLSEMMNEBFGVELDSTLLFNYPVIE	: : :	856
QY	65		-----	64
DB	857	IVDFVAAQISGKELHRKNLNRSGIDSYSDIWVGMSCSPFGSGSTPGEFWSLLQSGVDA	: : :	916
QY	65		-----QLETTISTK-----	73
DB	917	IGEIPKRWNIDEYVSTDLNKMVYKEGGFIENIDQFASFPKISHAEARSMDPOORI	: : :	976
QY	74		-----	73
DB	977	FLEKSFEALKDVGTYTIDTLKRNISVPAGCCSNWDTHVCKSEANSNIGTYAATSHAASII	: : :	1036
QY	74		-----	73
DB	1037	ANRVSYTLGLAGSSVTVDACSASIVLHSLQEIETQCEAAVVGGINLMSPOITVAL	: : :	1096
QY	74		-----	73
DB	1097	CKARLMSMDCRCKSFDAANGVVRGEGVGLVLTMTMENANNGEKIYSVIKGSSVNHNGRS	: : :	1156

QY	74	-----	73
DB	1157	ASLTAPNGISQONVINGSALEYAGIRPNDIGYIEAHGTGTSIGDPIEVSALKSIFSRTRDS	1216
QY	74	-----	73
DB	1217	GKPLIIGAVKTNIGHLEGAAGMAGIFKVLISLINDTVPNPLHFKNINPHINLKGPPVILP	1276
QY	74	-----	73
DB	1277	TSNIALQTYDGSKTCAGVSARFGGTNAHAIFEKPNLSIDLPIKTVKIETNKARNNPVLF	1336
QY	74	-----	73
DB	1337	VFGQGFQHVNMKYPEREIVYREWFUKCKIIDPFLKISLKLVSVPETKRMKRLBLN	1396
QY	74	-----	73
DB	1397	KILQBEYAQCALFAVQVSLAKLLESKGIYPCAVIGISDGLVASVYCGSINLEDGLKLC	1456
QY	74	-----YNVSK-----	78
DB	1457	MLRASLLSEFSAGSGKMAFPVSKEDAEMALADVDPQHNNRVISIFASISSNQVLLSGDBE	1516
QY	79	-----	90
DB	1517	KLETVPFILREYPRANAHSIDAKPLRTKYSGFTGNGMGLIEPISNLMENIALETMPIM	1576
QY	91	-----POY-----	94
DB	1577	ISGTSGYVVEELTSPKFWGQNLNKLKISKALTALINGASTILDMPISRSEFEISEI	1636
QY	95	-----	94
DB	1637	FNPNIDKDTIYSLPBDTNLESYEPNKNKMIKVVSILLKNPVNVEDACDQIINSNGS	1696
QY	95	-----	94
DB	1697	TYRESFPWNSFKHSILGEQKSIDRSVKIVSTLNKSIKLFSDHKVYETIIVPGAGLIDL	1756
QY	95	-----ILAG-----	98
DB	1757	AAASIIIVSMDDSILOSNALFGESIAVKLNVLFEKPVVNRTSLSSECTEIVINOLK	1816
QY	99	-----	98
DB	1817	SYNNLANDQKTNLICNIESDGSISIOYCNIDNQEDFDYADNSMIDCFCKAKISEGEFV	1876
QY	99	-----	98
DB	1877	LSEDLAQLKNNTIKEDPKIMYEYEBEIGLYKKFRVVKIYRNETCSIALGKIYLPET	1936
QY	99	-----PI-----	100
DB	1937	CKFTFSGFYFHPAILDGAHVAGSLLYKQKSPKQAMVPVSEIEIEIKYVNNQMIW	1996
QY	101	-----QWYSI-----	105
DB	1997	AAAYLSEATQNSIFDLVDFDSEGCIGKLEKVSRLQPNKNSIRPTVSHNEILLWTEWE	2056
QY	106	-----	105
DB	2057	KSDVLKFSKNISKESDPLKIALIHSGICNISEKODIQLPKANISFVNLDKIESNIBEL	2116
QY	106	-----	105
DB	2117	ERLIDITSLDSLVELSPVELALSAKIDCEKISIDILEYVMKICKLYLKLOAKPTKQEQ	2176
QY	106	-----	105
DB	2177	KINIPNFWIITTSQNLIYNINENICIPNNSGLWGLAKSANLEISSLLTNFSQPIKSV	2236
QY	106	-----	105

[illegible]

Db	3317	NGPAQTCVIMALNOSKLNSSDIDYETHGTGTALGDPIELGALKSVFGRNADYKRS	TPL	3376
Qy	164	-----	-----	163
Db	3377	VLGALKSNIGHLEGAAGIAGLIKVLVLKHETAIRIAHLTKINPHLDLPNVII	PQKTV	3436
Qy	164	-----	-----	168
Db	3437	SLKCNKNLIGVSGFGGCGNTHIVIVESTEKNSEDIAADFSKNHILFVFTGQGSQYI	3496	
Qy	169	-----	-----	170
Db	3497	NMCKGIYETEPFIRENMNLCNEIVSKILNISLLDIYPIDGNKDAEFKAMNMLNDRYAO	3556	
Qy	171	-----	-----	170
Db	3557	PAIFAVEYSLAQLWISKIKPDSVLGHSGLGEFAAATISQVMTTIEDALTIVTHRASIMAST	3616	
Qy	171	-----	-----	170
Db	3617	PALDGMVACRLTESQVDDTIKRFGLSEASLAUVNGQKSVTISGKESVYKILEYNNVG	3676	
Qy	171	-----	-----	170
Db	3677	SRFRDLVSHAFPLVSEASDKFLENIELKKTVEFISTVTGKKEISISTAKYWS	3736	
Qy	171	-----	-----	170
Db	3737	KHILNTVLSDAVHAIESNSRMTFIBITSKPIQLLKALIPVNSSHISVKCTCKLVN	3796	
Qy	171	-----	-----	180
Db	3797	SEIDPIQLEBQFEGSENSFTNHANNLDDSWRFKLKRRQISWDSISHPIIAPLHDP	3856	
Qy	181	-----	-----	180
Db	3857	NDIEWKENTISISKSNARELVEKLEFTCMISPELLDFDNHKVLQGSILPGAAFVDFMAT	3916	
Qy	181	-----	-----	184
Db	3917	VALNVTKSQFTGLVIGAMPDWIQLKGIFFRNPFIMSSAKQYKYLNTQDSDQYSDENSS	3976	
Qy	185	-----	-----	184
Db	3977	RSDPNIIMSGNKNCOISIESLKMDEETWVZATCDEVYVSSIEAQTKSLCIISEWPSS	4036	
Qy	185	-----	-----	184
Db	4037	ELISNKKVQEDIEYKMSIAGLOYGRFKTKEMWKVTSNVAVCIIIOSELFTRKCAESP	4096	
Qy	185	-----	-----	184
Db	4097	DNLAVYESIMNERGFTIHTPLLDGVLHMCASILGQSDKETFNELHKEATWVPISINKCL	4156	
Qy	185	-----	-----	184
Db	4157	ITSKIDTSKYIKEKWTAFVQLKSSDKOSAIVNVALKSMGVPPIAVLLGVSLRSVKNGVI	4216	
Qy	185	-----	-----	199
Db	4217	STHTIKHVIPNELLWRIDWDEVIESITNSEIVSLPTKVCNCONTENESVQGINANLVEKRK	4276	
Qy	200	-----	-----	203
Db	4277	VLLTNSCNISKELLYSEYFTADIVCEASKLSLAEIATIIYGRDQELPPSPSNEAKYDL	4336	
Qy	204	-----	-----	203
Db	4337	ICLHIFNSDKNAIVSLLPVLCKILRSMFENKERLIPRVRIITENI FNVDCMNLFR	4396	
Qy	204	-----	-----	214
Db	4397	NSGVAPFIKARQELFSDOSVSLGLDIDFABDLNKAIIILSFVKDILCFKKREDITSN	4456	

QY 215 -----AQR-----NCTHSFYLVN----- 227
Db 4457 GTLTESQRLFLNTEYEPELVKKGKEDNCEAHYRIITPKVLSLGERAAIGSKLNL 4516
QY 228 ----- 227
Db 4517 SSRGAINNLELLPLSNEERKCPESANTVIRVRSIGLNFDRVLNMGLYPGDPGPGDCS 4576
QY 228 ----- 227
Db 4577 GTVAVGEGVKHIVGDNVFIAPGLCKTYVTDSNLLCKIPKGTPEQAAALPVVATTV 4636
QY 228 -----AMSRNLFR 235
Db 4637 EYSLRDIANIKKGDIVLVHAVTGGVGLMVVQYCKAIGAKVGTAGSKEVYALSNGVER 4696
QY 236 V----- 236
Db 4697 VSSSRNADKFKEDMKDLEGKVDIVINSIEDIPNLSNLLKKGCFVELKRGKIWTEEM 4756
QY 237 -----PKYING--TKLKN----- 247
Db 4757 KEKRPDIKYCVADVMMEEDEPSWFGMLTRIKNLVEEGIIIESLPLKYVDMRGSDENGID 4816
QY 248 ----- 247
Db 4817 AFYMQRAQHIGKVIKIPTPPKYLELDTENHKLNNVSEKEDRHNHVEYGIHVITGGIGG 4876
QY 248 ----- 247
Db 4877 IGRIIWKMLBEGVRKIAILSRSANNDSLRDIPEIKOYLNTDIIQIECIKCDVSILSQVE 4936
QY 248 -----TMRKLKKEQAP----- 258
Db 4937 SAFREILVKFGETNQIHGI FRAAGILMDGATASQTMEMWESVYAPKYVGAWNLHECCBKF 4996
QY 259 ----- 258
Db 4997 DLNKLKHVMFSSVASLGNFGQNTYSAANCLDSLVEYRRNKGMCOTSIQGWPIBQG 5056
QY 259 -----VKEQFE----- 264
Db 5057 MAMNIKOHLEKVGHMGISNEMGIRVINDVILYQNTDLDKNSLGPYPVITCOSFNWVYMR 5116
QY 265 ----- 264
Db 5117 QLYNSKYPSPFKDVVASMSTGEVKNIGMDFRGMSEVYVYKDSVHNIVEIQIIGQSS 5176
QY 265 ----- 264
Db 5177 TLDLTPFSELGMDLSAIELRNALISRFGIKISSTALFDYPTLRSIIQHLTDSLIVSET 5236
QY 265 ----- 264
Db 5237 RDAFSDLGMSISEREYVAIIIGVSCRPLGRVKSFNLDYEIVLLKSKDCITBPLDRWPF 5296
QY 265 ----- 264
Db 5297 LYDPDPDNRSKSVYNOQAFITDIDL PNTPFGLSSTEATNVDPQKYMLEVAYEAMVNS 5356
QY 265 -----KKAKKT----- 270
Db 5357 GITKRLSITSIKTAKTGDDPVLGMGVFIGCCNTDWHFLQAKTFENFTSYTSGSGGSSLVS 5416
QY 271 ----- 270
Db 5417 NRVSYSFGRGFSQITDTACSSSLVAMDAALKLLKSGSHTALVGGVNLMLSPHLFVAF 5476
QY 271 ----- 270
Db 5477 RARMLSPDCRCKTFDASANGYRGECCGAILLTRSEIGGKIRPIAYVLGSATNHNRS 5536

QY 271 ----- 270
Db 5537 ASLTAPNGPSQTEVIQNALRNARLTPNDIDYLESHGTGTPLGDPIEFGALKTVFGKKXI 5596
QY 271 ----- 270
Db 5597 KRSQPLILGALKTNIGHLEGAAGVGLLKLILVLKNRIAPKILHFKOKLNPHTDNEFDVE 5656
QY 271 -----QSTTTPY--FSYT-- 281
Db 5657 PPTELKPIISNKSXLVGVSSFGGCSNVHLESASANELNTESSQKNYPYLFIAYTG 5716
QY 282 -----TSA 285
Db 5717 QGCQYINMKQLYDTEEVFRNEMDKNELLYPILGRSLISTIYCEKTMESKLIHTSIS 5776
QY 286 LNVITNTVYS-----ITTAARRV----- 303
Db 5777 QPAIFSIYSUTKLWESKGI VPSAVMGHSLGEFAAAVTSGILCLKDAITLVARRAEIMAS 5836
QY 304 -----STSTIA----- 309
Db 5837 LPQNDGIMVACRVSEDQVLD SINKLNKDSAAVAALNGPKSVTISGSRSSVMQVLNLMG 5896
QY 310 -----YRDPSSPMKSI----- 320
Db 5897 EDRYKQLDVSHAFPLVSSABEKFVKVNIQFPEPKLTFPSSVTGNIESKLIENKYW 5956
QY 321 -----MATQLRDLA----- 329
Db 5957 AQHIIHVVRYDALETAISTDEENIAIEIGPKSVLTSMGKQAVGGIOLVOSSNENISNN 6016
QY 330 --TWV----- 332
Db 6017 ELTWIPTIDNSKNTIDNDALKAAREYIKYISBVCVHKWNHRSRPHWADLTSVHPILGK 6076
QY 333 -----YTT----- 335
Db 6077 FYSSISDENLAYSTDMLSKSVFKNI VEDHVI FDTNIIIPAAALIEASACAAPLQRSEKF 6136
QY 336 ----- 335
Db 6137 PSKYVENDSMDIEKPLVSNODETTSEMEVFVAKGAIQIRKCI SEAFYEFEFSGPOII 6196
QY 336 ----- 335
Db 6197 ANDTVFSNRIYHHASISNSRILSKDEVDKQCPNLDVLYLEKYSKDTMEVPIDNIYQM 6256
QY 336 --LRYQN----- 341
Db 6257 NEIGLQYQKNYQCLKT VKVKIDNSDEKKIKSGSNQINEVFCYVXRPI SDPDLTELRIH 6316
QY 342 ----- 341
Db 6317 PGMLDSVFQASSVLFA SDELLEGKDTTKNKRGHVAMAPIGTCKFCYGRIOHNSDWMGLVK 6376
QY 342 ----- 341
Db 6377 LNRPEKKGNI AFIDVTLWDSKTNILAAHFQDILLKGFPSNAASLLSQEVRIPNQPLWRTYN 6436
QY 342 ----- 341
Db 6437 KKI VDRNTYIENSSTPFEQENKEDQKBFNSEDSDHDNQIOKILIFGIPTVIKELICSQMD 6496
QY 342 ----- 341
Db 6497 NKYKISTTTFLEEDLTIDLIDOKI QEENWSAMFYLGGLMDNLSKQVMGDILKISQVI 6556
QY 342 -----PF-----CE----- 345
Db 6557 STISVSITTAFFVCLT SFNDCEDGNIPRHSGIKGVKTARVELENVIGKISINISHIECN 6616
QY 346 ----- 345

Db 6617 SSLVNDPBLTMTISYIASIYEGKIEBITLERDLTISESIVAPRLSINFPVFGAIKL 6676
QY 346 ----- 345
Db 6677 TMSRGAIGNLKPKPOQSERVNPANTVEIRVSIGLPRDVLNVMGLYPGDPGGD 6736
QY 346 ----- 345
Db 6737 CSGTVVAGGVKHKVGDNVFGIAPGCLKTYVITDSNLCKIPKGTPEQAALPVVAT 6796
QY 346 ----- 345
Db 6797 TGEYSLRDIANKGDIVLVHAVTGGVGLMVQYCKAIGAKVYGTAGSEKVEYALNGV 6856
QY 346 ----PSENRTAVSEFMKQTH----- 361
Db 6857 ERVSSSRNADKFKEDMKDLGKVDIVINSLIEDFIPNSMLLKGGCFVELGKRGIWTEE 6916
QY 362 ----- 361
Db 6917 EMKEKRPDIKVCVAVDVMMEEDPSWFGMLTRIKNLVEEGIIIESLPLKVMRSGDENG 6976
QY 362 ----VLIRNETPY----- 370
Db 6977 IDAFRYMORAQHTGKVIKIPTPKYLELDTENHKLNVGSEKEDRHNNYEYGIHVTGGL 7036
QY 371 ----- 370
Db 7037 GGIGRIIWKMLEGVKRTAILRSANNDLSRDIPEIKDYLTNDIIQIECIKCVSILSQ 7096
QY 371 ----TIV----- 373
Db 7097 VESAPREILVKFGETNQIHGIFHAAGILMDGALASQTMEMMESVYAPKVYGANWLEHCE 7156
QY 374 ----- 373
Db 7157 KPELNKLNKHFVSSVASLLGNFGQTNYSAA NSCLDSLVEYRRNKGCGTSIQGWPWIE 7216
QY 374 ----- 373
Db 7217 QGMAANLKOHLKVMGYGISELGIKRVLNDILCFVSKSSVIGCOSLKWDTFMRRYNTIP 7276
QY 374 ----GTLDMSSL----- 381
Db 7277 TFLSEVDYSSQGRSQGRVDISKLSDEELVWVIAQVAAQCASPTLPSPDTVLLDGLDSL 7336
QY 382 ----YNETMFVE-----NKTASDSNK----- 399
Db 7337 GAVEFRNSVLEMTGVKLPQTLVPENPTIYATSMYVRDQNSGNTKSDSKHQSOTSNEQAK 7396
QY 400 ----- 399
Db 7397 MTIEQWLLSSLKQSERIYIAEAFKKYQTI SNLTETDIISALEELGVENNOYDILLYV 7456
QY 400 ----TTPTSps----- 406
Db 7457 SNNELISEEENSKONDSNSKLEQELIFDPIADVEIKNSMKLDLDGISTTDPSEFK 7516
QY 407 ----- 406
Db 7517 TALLTGVTVGYRILLVVKLIEBFKNMQIVCLVRASSEKGLERININVEERAVWNPTYAS 7576
QY 407 ----- 406
Db 7577 RIIVECGNFEYELGLSEERYELCKEIDVVYHIGDVLNLLSNYKRLKNTLSLVGIN 7636
QY 407 ----- 406
Db 7637 FCTIKLKLHFSSTLGQFPAPFAVFTREPENAVVKEGPESTREMSRLFPPTROQYPWS 7696
QY 407 ----- 406

Db 7697 KWAEBQILETAHQOGLPLSIYELPNTYIASDTGYTNKTDYATALLIASILEGMRPFGSST 7756
QY 407 ----- 406
Db 7757 ABLTPVNTICEIIISASKKKERKWRNLLDTRILGKKHVEAWASQLGLGNYKGVAIDEF 7816
QY 407 -----MGF 409
Db 7817 FSAIKSRGPSPFKFVPLMQVWRHYWFDNTERTOPFPVDTSNVSEDMPEISWPLEITF 7876
QY 410 QRTFT----- 414
Db 7877 QNSFLYCAKRGYFPSPNSSSISFTPIKCYSDALNEIKAELSLRNVFNQNKYDFICRKMOK 7936
QY 415 --DPLWDYL-----DSL----- 425
Db 7937 NADSIYDSLKKVSLNFCNCAKSYIILLKQSVQVYVQESLNDLSILEKFMQMDCYFITGTS 7996
QY 426 -----F 426
Db 7997 FEIELELRVIRKLLNNYNELKFVDIACPYXPKNSRRLILEQLKIATSIPLKRFPAVD 8056
QY 427 LDEI-----RNP----- 433
Db 8057 LDEILTDSDILLEMFPAPFTPLIYKGYFEEYKKNKITCSEVMEAYSSFQNYIIQTLI 8116
QY 434 -----SLRSPTY 440
Db 8117 KKNYNTVEPIIFSSPFFHLPFVKEILOIFENSKIIVVDNKPNDSTIQGKLSMWKSLKS-KY 8175
QY 441 VNLTPPS-----HRAVN-----L 454
Db 8176 TNLDSVEENSEVYLLDILEMKHRFVINEPSIDKNRLIYLKDFGEIHNEKISVOLQDFC 8235
QY 455 STS 457
Db 8236 STS 8238
RESULT 6
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C;Species: Drosophila melanogaster
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C;Accession: T13931
P;Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A;Title: Structure of the Drosophila projectin protein: isoforms and implication for pro
A;Reference number: Z17815; MUID:98300339; PMID:9636710
A;Accession: T13931
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-6658 <DAL>
A;Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:G3337430; PID:G3337431; PIDN:AAC2
C;Genetics:
A;Gene: projectin
A;Cross-references: FlyBase:FBgn0005666
A;Map position: 4
A;Note: intron positions not resolved (incomplete sequence)
C;Keywords: muscle
Query Match 20.4%; Score 500; DB 2; Length 6658;
Best Local Similarity 3.1%; Pred. No. 0.28; Mismatches 130; Indels 6187; Gaps 68;
Matches 208; Conservative 98;
QY 5 EMVVR-----DVPKMFVLISISFLVSPINCKVMKALYNRPW---RGL----- 45
Db 1 EFRVAVNRAGPGEPSDKTK-----SIIAKCRFVKPFIVCEGLKNVTV 43
QY 46 -----VLSKIKY 53
Db 44 KKGQIRFDIKVDGEPPEPAATWKTGDNLKNFDNORICLDQLERNSSITIKKSVRDTGKY 103

54 QY KL-----55
104 Db KLVLNSSGTTESEAQQVVLDRPLPPGPFEPBEIRASHIKWKWRPDDDDGGCISGYAL 163
56 QY-----55
164 Db BRMBDETGRWTPAGEVGNETSFDKGLTPNKKYKPRVKAINKEGESEPLETFDASCCKNP 223
56 QY-----55
224 Db YDPPSPSQPVIDDYDNKSVLLKWKRPDSDGCRPITHYIVEIKDKFAPMSWSEVAKTDDPN 283
56 QY-----55
284 Db PECNVEGLKERWYQFRVRAVNKAGSPSPQTDNHLCKHKNLKPQIDRSTFRKRTIKSG 343
56 QY -DOLKLE-----ILR-----64
344 Db RTHKWSVDVLGEPIPELHWSRDDIPLTNGDRIKIENVDYHTDFSITNVLKDSGFYTLK 403
65 QY-----64
404 Db AENRNGIDRETVELVWLKPSKPGPLAVSDVTASGSKLQWKPEDEVGGVPIKEYVVEKM 463
65 QY-----64
464 Db DTATGKWRVGRSPGEKEPPSFDVTGLSLGSEYMFVRSVAVNEEGDAEPLTLVGWVAKDP 523
65 QY-----Q 65
524 Db PDEPNKGTPEVTDYDNQISLSKWAAPNNDGAGPIQKYIIEKKNKNTWEKALEIPEQDQ 583
66 QY LETTIS-----TKY-----74
584 Db LEATVAGLQEGYQFRVIAVNKAGLSPSPDASVPQIVKYKKLKPDRDRSLKPLLRAG 643
75 QY -----NVSKQP-----VKULTWNTFEP-----QYY 94
644 Db KPIRYDVNVGRGEPAPVITWYQNDKELPEELPSSSEIKNIPYNTKISIIETVRKHTGIYK 703
95 QY ILA-----97
704 Db IIAVNEHQDEATVEVNI LAPSPKRGPLDVKQVTKDSCPKWKPKBEDDGGKPI SAYOVE 763
98 QY-----97
764 Db KFDKKGRWVPLGRTSANDTEFDVKGLOEGHEYQFRVKAINEEGESDPLSDSDSIIAKNP 823
98 QY -----GPIONYSI-----TYLWPDFYSTQ-----116
824 Db YDAASKPGTPNIVDYNHVMVKLWEAPRSDGAGPISGYIIEKKDKFSPINDEILSTNTSV 883
117 QY-----116
884 Db PEATVEGLVEGNIYQFRVRAVNKAGSPSPDATEPHLAKPRNLKPYINRDKMKPIKVRAG 943
117 QY-----116
944 Db QPVKFDVDVKGEPAAPSLTWFLKETELTSTGQVRLENIDYNTKLTLLDTRKQSQYKLRA 1003
117 QY -----LRKPAK-----YVYSQY-----128
1004 Db ENINGVDBAVVEVIILOKPSKEGPELVSDIHKEGCKLWKRPKDDGIPITGVVIEKMD 1063
129 QY -----NH-----TAKT-----134
1064 Db TATGKWPAQSDPEKYDIEIKGLDPNHRVQFRVKAINEEGESEPLETESAITAKNPFV 1123
135 QY -----ITERPP-----140
1124 Db SAPPLPELEDWDEHHVKLWEPPIRDGGSPITWYIIEVMDKDPGEFVKA VETDPSVCKG 1183
141 QY-----140

1184 Db VVKLEBGOQYKFRVRAVNKAGSPDSEQTNHVAKPRFLKPHIDRVNLKPVIVKTGLSI 1243
141 QY-----140
1244 Db SLDINIRGEPAPKVEWFFNNSSVTSDEHSVKIDNVYNTKFFVMRAQRSQSKYIIKATN 1303
141 QY-----140
1304 Db EVGDEAELEVTVLGKPKPKPLQVNDITKHSCKLWKWKDDGGSPDIYVEIEKLDPH 1363
141 QY -PCGR-----144
1364 Db TGQMLPCGSTPEPAKVIGLHGBKAYFRVRAVNKEGESEDELETEKPIIAKNPVDEPDRP 1423
145 QY -----VPSMTC-----150
1424 Db GKPEPTWMDKDFVDLAWDPKNDGAGPIQKYVIQMRDKSRAWVDSATVPDCKNGVTG 1483
151 QY -----LSEMLN-----156
1484 Db VEEGHEYEPRIVAVNKAGSPSDSVKSIAKPRFLKPHIDRKNLQKKIMRSGQMLHIDA 1543
157 QY -----VSKNDTG-----164
1544 Db LIAEPPAKVTWYNTKTEIKTSDHIKIENEDYKTFIMPVKVRADRGYIIVTAKNDSGSD 1603
165 QY-----164
1604 Db TVEVEVLCKPKPKPLAVSNVTAETLHLKWKPEDDGDPIDIEQVLVERMDTETGRWV 1663
165 QY-----164
1664 Db PVLTTKTPEADVTGLTEGKEYLFRVKAINEEGESEPLVTDIPTKAKNPDAADTPGKPOI 1723
165 QY-----164
1724 Db VMSGNHCDLWRAPEDDGGASITGYIVERKDPNTGKWKQKALETSTPDCKARVNDLIAGN 1783
165 QY-----164
1784 Db KYQPRIMAVNKAGSKSPSBDQMTAKDRFAPPKIDRTNIKDITSKAGQHIRFDIKVSGE 1843
165 QY-----164
1844 Db PPATKVLHKNARLENDSDSNYIDMESYRTKLTVPISKRPHSGKYTLKAENESGRDEASF 1903
165 QY -EQCG-----169
1904 Db EVIULDKPGPEGLRVTDVHKEGCKLWNAPLDDGGLPIDHYIIEKMDVESGRWLPSGR 1963
170 QY-----169
1964 Db FKESFABLNNLSPSHYKFRVLAVENTEGESEPLTGEQSVIAKNPFDEPGKGTPEAVDWD 2023
170 QY-----169
2024 Db KDHVDLWVRPPINDGSSPITGVVVEKREKGTOKWIKGTEITIPCLGEECKATVPTLNENC 2083
170 QY -----NFTTFNPMFF-----179
2084 Db EYBFRVKAINAAGPEPSDASKPIITKPRKLAPTLDPKTNIIRTYNFKSGSPIFLDINIS 2143
180 QY -----NVPRWNTKLYV-----190
2144 Db GEPAPDVTWNNKNSVOTTSFHIENLP-YNTK-YINNPERKDTGLYKISAHNPFYQDQ 2201
191 QY -----GPTKV-----195
2202 Db VEFQINIITKPGEGGLEVEVHVKDCKLWKWKPKDDGGEPVESYLVERKDPDTGIWLP 2261
196 QY -NVD-----SOTIYFLG-----206
|||

Db 2262 VGRSDGPEYNVDGLVPGHDKYKFRVKAHVKEGSEPLETIGSIIAKDPFSVPTKPGVPEPT 2321
QY 207 ----- 206
Db 2322 DWTANKVELWPEPASDGGSPIOGYIVEVKOKYSPLEWEKALETNSPTTATVQGLIEGNE 2381
QY 207 ----- 206
Db 2382 YQFRVVALNKNLSEPSDSKIFTAKRYIAPKIDRNLRLNLTSSGTALKLDANITQOP 2441
QY 207 ----- 228
Db 2442 APKVEWKLNVHLOGKNVTIETPDYTKLVIRPPQRTDSGE- YLVATNTSGKDSVLVN 2500
QY 229 ----- 228
Db 2501 VVITDKPPNGPLQISDVHKEGCHLKWKRPSHDGGTPIEYFQIDKLEPETGCWIPSCRS 2560
QY 229 ----- 237
Db 2561 TAPQVDVTGLSPGNEYKFRVSAVNAEGESQPLVGDESIVARNPPDEPGKPNLKATDWDK 2620
QY 238 ----- 239
Db 2621 DHVDLAWTPPIDGGSPISCIYIEKQDKYKWERALDVPADQCKATIPDLVEGQTYKFRV 2680
QY 240 ----- 249
Db 2681 SAVNAAGTGEPSDSTPPIIAKARNKPPIIDRSSLVEVRIKAGQSFTFDCKVSGEPAPQTK 2740
QY 250 ----- 256
Db 2741 WLLKKKEVYXDNVKNVTVDYNTKLKVNSTRSDSGIYTVFAENANGEDSADVKVTVIDK 2800
QY 257 ----- 259
Db 2801 PAPPNGPLKVDINESECTLHWNPPDDGGQPIDNYVVGKLDETTGRWMTAGTDGPVTA 2860
QY 260 ----- 286
Db 2861 LKVGGLTPGHKYKFRVRAKNRGTSSEP--LTTAQAIIAKNPFDVPTKPGTPTKDFDKE 2917
QY 287 ----- 293
Db 2918 FVDLEWTRPEADGGSPITGYVVEKDKFSPDWEKCAEISDDITIAHVFDLLIEGLKYERV 2977
QY 294 ----- 293
Db 2978 RAVNKAQGPSDATETHVARPKNTPPKIDRNFMSDIKIKAGNVPEFDVPTGPELPFSKD 3037
QY 294 ----- 304
Db 3038 WTHEGNMIINTDRVKISNFDDRTKIRILUSATSDTGVYTLT--ARNINGTDRHNKVKVILD 3095
QY 305 ----- 304
Db 3096 APSVPEPALRNGDVSKNSIVLRWRPPKDDGGSEITHYVVEKMDNEAMRWVPVGDCTDEI 3155
QY 305 ----- 304
Db 3156 RADNLIENHDYSFRVAVNKGQSQPLTTSQPIITAKDPYSHDPKPGQPQATDWGKHFDL 3215
QY 305 ----- 304
Db 3216 EWSTPKRDGGAPISYIIIEKPKQWERAADVLDNCKAHVPELTNGGEYEFVRIAVNR 3275
QY 305 ----- 333
Db 3276 GGPSDPSDPSSTIICKRFLAPPFDKSLNLDITVHAGNGLGWLTPLEASPRPLITWLYNG 3335
QY 334 ----- 333
Db 3336 KEIGSNRSGDSLQFQNELTFEIVSVLSRABGRYTLIILKNEHGSFDASAHATVLDPRSPPK 3395

QY 334 ----- 337
Db 3396 GPLDITKITRDGCHLTWNVPDDGGSPILHVIIEKMDLSRSTWSDAGMSTHIVHDVTRLV 3455
QY 338 YR----- 346
Db 3456 HRKEYLFRVKAHVNAIGESDPLEAVNTIIAKNEFDEPDAPGKLIITDWRDHDIDLOWAVPK 3515
QY 347 SRNRTAVSEFM----- 360
Db 3516 SDGGAPISEYIIQKKEKSPYTNVRHVPSNKNNTIIPELTEGQEYEFVRIAVNAQGOSE 3575
QY 361 ----- 360
Db 3576 PSEPSDMIKRPRYLPPKIITPLNEVRIKCGLIIFHTDIIHFIGEPAPEATWTLSNPLLSN 3635
QY 361 ----- 366
Db 3636 DRSTITSIGHHSVVHTVNCQSDSGSIYHLLLRNSSGIDEGSFELVVLDRPGPPSGPMEYE 3695
QY 367 ----- 366
Db 3696 EITANSVTISWKKPPKONGSGSEISSIVIEKRLDTHGGWVPAVNYVSAKYNHAVVPRLLG 3755
QY 367 ----- 366
Db 3756 TMYELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAPCKBELTSDKNHITIKWKQPIGN 3815
QY 367 ----- 366
Db 3816 GGSPIIGVDIERRDVNTGRMIKINGQVPVTAEYQDDRVTSNHQYQYRISAVNAAGNGKTS 3875
QY 367 ----- 366
Db 3876 EPSAIFNARPLEKREFYFDGLIGRIKVRAGEPVNLNIPISGAPTPTIEWKRGDKLEE 3935
QY 367 ----- 366
Db 3936 GKRISETNSERTLFRIDDSNRDRSGKYTVTAANEFGKDTADIEVIVVDKPSPEGPUSY 3995
QY 367 ----- 366
Db 3996 TETAPDHI SLHWYSPKDDGGSDITGYIIEFTFEGVDWKPVPGTCPTNTFTVKNLVGKK 4055
QY 367 ----- 366
Db 4056 YVFRIRAEITYGASEALEGKPVLVKSPDPGAPSQPTISAYTPNSANLEWHPPDDCGK 4115
QY 367 ----- 372
Db 4116 PITGYIVERRERGGEWIKCNYPPTNTSYTVSNLRDGARVEFRVLAVNEAGPHSPKPSD 4175
QY 373 ----- 372
Db 4176 PMTAEHQRYRDPPEPPKPKDRITRNGVTLSWRPRTDGSRIKGYVYEMRPNKGDKWKT 4235
QY 373 ----- 372
Db 4236 NDIPINSTVTVPSLKEGEYSFRVVAENEVRSDPSKPSQPIITIEBQPNKPCMELOKVR 4295
QY 373 ----- 372
Db 4296 DIVCRAGDDFSIHVPYLAFFKPNAPWYNDNMLDDNNRVHKLHTDDAASVVVKNKSKRDS 4355
QY 373 ----- 372
Db 4356 GQYRLQKDTSGFDTATINVRVLDPRSPPTRLRADVFGDSLTYLWNPNDGGSAIQNY 4415
QY 373 ----- 374
Db 4416 IIEKKEARSWSKVSFCTVLVFRIRNLVLNKEYDFRVAENKYGOSDPANTSEPIILAR 4475

QY 375 ----- 374
Db 4476 HPPDIPNTPGPHGIDSTEDSITIAWTKPKHGGSPITGYIIEKRLSDDDKWTAKVAHALC 4535
QY 375 ----- TLDS- 379
Db 4536 PDLCKIPNLIAENAEYFRVAANVAGOSAYSGSDDLFCRRPPHAPKITSDLSIPDMTV 4595
QY 380 ----- SLYNET 386
Db 4596 IAGDEFRTVPYHASPRTASWSLNGLEVI PGERIKPDSNDYASWYKSAKRDGTGTYT 4655
QY 387 -MFVENK---TAS- 395
Db 4656 ITLTNNKGSOTASCHVTVDRLPQPQPLNAYDITPDTCTLAWKTLPLDDGGSPITNYVVE 4715
QY 396 ----- 395
Db 4716 KLDNSGWSVKISSFVRNTHYDVMGLEPHYKYNFRVRAENQYGLSDPLDIIEPMVAKHQFT 4775
QY 396 ----- DSNKTT 401
Db 4776 VPDPGQPKVIDWSGNVTLIWTRPLSDGSRIOGYQIEXRDILNDSSNAYDYIIKGTK 4835
QY 402 ----- PTSPSGFQ- RTFID 415
Db 4836 YQLYNLINGSYEFRIKAKNAAGLSKPSPSLRFLKKGKFTVPSPGAPQVRVGKNYVD 4895
QY 416 ----- 415
Db 4896 LWKELPUDGSRITGYIIRRDIGGAVWVKCNDYNVLDTBYTNLIEMGDYFRVPAV 4955
QY 416 ----- 415
Db 4956 NSAGRSBSLCTMPIKCEVLGKKPDWITRLQKVAPPKDYTLQCAASKPSPATRWL 5015
QY 416 ----- 415
Db 5016 RKGKEIQWNGGRMTCDSKGVFLHISNVQTGDDGYTCEAMNSLGFVNTSGYLKIGSPI 5075
QY 416 ----- 415
Db 5076 INRCPELKLPEGONSKIKIPYSGDQPLTVILKQNEVISDSNDTHVKVNI PDDYVAIY 5135
QY 416 ----- 415
Db 5136 IANIVKSDGGPYQIEFTNESGSATGEFVYHITGHPAPTGPWGISYINKNSCMLNWRPPY 5195
QY 416 ----- 415
Db 5196 YDGLKVSHVYTERKDVSSPHWITVSSCTKDTAFNVQGLIENQBYIFRVMAVNEGMGPP 5255
QY 416 ----- 415
Db 5256 LEGLNPIRAXDPIDPPSPPAVLKSLRSEETLCNPEWKKPESDGAHIQGYWIDKREVGSK 5315
QY 416 ----- 415
Db 5316 HMAGVNATICANQINCINLIEGRQYBRIIPAQNVAGLSTAIVRFTSQDNRSSTGGFASI 5375
QY 416 ----- 415
Db 5376 DCETATHANCIONHNAQFTCTIPGVKPTISWYKGAREISNGAGITCTQKVTSLPKYVDV 5435
QY 416 ----- 415
Db 5436 FGEDAEYVCRAVNKAGAKSTRATLAINTAPKLVNPPFRDPTAYFDKGENVVIKIPTGL 5495
QY 416 ----- 415
Db 5496 PKPRIHWKLENTESGHHYTVVEKHAVALIIRDSHLSDGPPYRTAENELGSDTAIIQ 5555
QY 416 ----- PLWD-----YL----- 421

RESULT 7

Db 5556 VOISDRPPRPFPLIESIGTESLSLWKAQVWDCSDITNYVVERRDDPLSWIRVGNTR 5615
QY 422 ----- 421
Db 5616 FTSMVAGSLTPGKEYDFRIPADNVYGRSDASDTSTLTKTESVKKKPIERKWEIDANGRK 5675
QY 422 ----- DSLLF- LDEF- 430
Db 5676 LRKADGPVKDYDSYVFDIYSKFPQPVBEISQQSVYDRYDILBEIGTGAFGVWVHRCRBS 5735
QY 431 ----- 430
Db 5736 TGNIFAAKPIPVSHSVEKDLIRREIDIMNQLHOKLINLHDAFEDDDDEMILILEPLSGGE 5795
QY 431 ----- 430
Db 5796 LPERITAEGYVMTAEVINYMRQICEGIRHMEQNIHLDIKPKENIMCQTRSSNTNVKLID 5855
QY 431 ----- 430
Db 5856 FGLATRLDPNEVKITTTGTAEFGAPEIWNREPVGFTDMWATGVSLLSGLSPFAGDN 5915
QY 431 ----- 430
Db 5916 DDQTLKNVACDMDFALESFKYISEBAKDFIRKLLVRNKEKRMATAHECLLHPWLTDGSHA 5975
QY 431 ----- 430
Db 5976 MKQBINRDRVLAVERKURRYVEDFERFLLPTGLRSLSEYSSRLKLLMEKYIHDVAFDRQA 6035
QY 431 ----- RNFSLSRPT- 439
Db 6036 APRFVIRPSSQFCYEGSVKPYCFCRCIAIATPTLTWSHNNIELRQSVKFMKYVGDYVFI 6095
QY 440 ----- YVNLTP- PEHRR- 450
Db 6096 INRVKPHDRGEYITRAENHYGSRBEVVFLNVQPLPKQEPRYRTSTPVRREPLPYTFWQ 6155
QY 451 ----- 450
Db 6156 ESEETAPSTFLLRPRVWQARDCTKLCCLSGKVPVNVWYKQGRSLSKYBYAMTHSDGV 6215
QY 451 ----- 450
Db 6216 VTMBEIDCKPSDSGNYSCKATNCHGTDETDCWIVEGEWVTPEQAQLAHNFLYSGDRKYI 6275
QY 451 ----- AVNLSTNS- 459
Db 6276 EQPIKAPLPITVTSRQYTSVQNTSBPQSGDKVNVSNSSSGISNKKKYASNLQAPGSP 6335
QY 460 ----- 459
Db 6336 SRSRSATKELLPLPDDSLMCKPFTKPLHDLTHDGEQLILTCVVGDPBPQISWSKNGK 6395
QY 460 ----- 459
Db 6396 SLSSDILDLRYKNGIATLATINEVPPEDEGVITCTATNSVGAVETKCKLTQPLDKNINK 6455
QY 460 ----- LW- 462
Db 6456 RKNVAGDNAPKIVSHLESFRVDRGDVNLACRIIQAQHFVWVWLHNKEIKPKSKDFQVTN 6515
QY 463 ----- 462
Db 6516 EANIYRLOIAEIPPEDGGTYTCEAFNDIGSEFSFCTINVTVPGETKQPSFVKPPTSVSV 6575
QY 463 ----- WLQ 465
Db 6576 LEGEGTTFECEIDSELNLVWLK 6598

Tl4593

syringomycin synthetase - Pseudomonas syringae pv. syringae
C:Species: Pseudomonas syringae pv. syringae
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: Tl4593
R:Guenzi, E.; Galli, G.; Grgurina, I.; Gross, D.C.; Grandi, G.
J. Biol. Chem. 273, 32857-32863, 1998
A:Title: Characterization of the syringomycin synthetase gene cluster. A link between pr
A:Reference number: Z18153; MUID:99047670; PMID:9830033
A:Accession: Tl4593
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-9376 <GUE>
A:Cross-references: UNIPROT:O85168; EMBL:AF047828; NID:g3510628; PID:g3510629; PIDN:AAOB

C:Genetics:
C:Gene: syrE

C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:441-896/Domain: acetate-CoA ligase homology <ACLI>
F:914-981/Domain: acyl carrier protein homology <ACP1>
F:1529-1984/Domain: acetate-CoA ligase homology <ACP1>
F:2002-2069/Domain: acyl carrier protein homology <ACLI2>
F:2613-3071/Domain: acetate-CoA ligase homology <ACP2>
F:3089-3156/Domain: acyl carrier protein homology <ACLI3>
F:3700-4158/Domain: acetate-CoA ligase homology <ACP3>
F:4176-4244/Domain: acyl carrier protein homology <ACLI4>
F:4768-5224/Domain: acyl carrier protein homology <ACP4>
F:5242-5310/Domain: acetate-CoA ligase homology <ACLI5>
F:5834-6280/Domain: acyl carrier protein homology <ACLI5>
F:6892-7352/Domain: acetate-CoA ligase homology <ACP6>
F:7370-7437/Domain: acyl carrier protein homology <ACLI7>
F:7370-7437/Domain: acetate-CoA ligase homology <ACP7>
F:7990-8440/Domain: acetate-CoA ligase homology <ACLI8>
F:8458-8526/Domain: acyl carrier protein homology <ACP8>
F:9015-9083/Domain: acyl carrier protein homology <ACP9>
F:946,2034,3121,4208,5274,6330/Binding site: phosphopantetheine (Ser) (covalent) #statue

Query Match 20.4%; Score 500; DB 2; Length 9376;
Best Local Similarity 2.5%; pred. No. 1;
Matches 232; Conservative 98; Mismatches 123; Indels 8785; Gaps 80;

QY 1 MGRKEWM-----VRDVPKM-----14
DB 1 MGGQRLSKAPFKLEVTDFEVDGTEQAIAHRESFRPPESVTGQWESLYRCGPQL 60
QY 15 -----FVLIS-----ISFL-----23
DB 61 HYWFNRLHLVADGIGAVLIAHAVSDAYSGLLEGKETLAEGPSYLSFLEEDRAYMASSRY 120
QY 24 -----23
DB 121 KRDLRFWEQSYAQLPPVLLQRRADFAGLTNVPAPSDQVQAMLPRALYNALYQFASCNL 180
QY 24 -----L 24
DB 181 SVAHVLGVISTYFCRTGVDBEIVGMPVHNRTNARQKATIGMFSVPINRGDPHETL 240
QY 25 VSFIN-----29
DB 241 VDLMMNIGTLQRLRTYRHQFPFAELNLSRLAQNQRHQLFDVSUSPESFGDVTFGSDP 300
QY 30 -----29
DB 301 ARVLMNDNGYEQTPMAIFVRDYHPSEDIHLDNFNTAYFTFEAAQRVQQRIFAMLEAVLN 360
QY 30 -----29
DB 361 QHDTPLAHFPLMSTSEHQALATFNDTAHPYPRDVLIIHQLEQQAAQRPDACAVRGDSGFL 420
QY 30 -----29
DB 421 LTYAELNQANQLAHLRIELGVDPTRVAVSLRGAEMVVALGILKAGAYVPIDPDLP 480
QY 30 -----29

DB 481 SARQAYMLEDSSPQAVLTTRDLSDNLPSADLPVLVLDGHNDRAQLARQQSVNPDAKALGL 540
QY 30 -----29
DB 541 QPNHLAYVLYTSGTGTGKGVNHLGVNRLNWARDAYQVNSQDRVLQKTPCQFDLSVM 600
QY 30 -----CKVMSKA-----36
DB 601 EFPFLTLTGAELVMAPPGHQDDPYLAQVMSDAGITLLHFVPSMLDVLEHSTRDPQL 660
QY 37 -----LV-----38
DB 661 RRVLCSEALPRALQRRFEQQLKGVELHNLGPTAAIDVTAWECRPTDQDGSVPICRPI 720
QY 39 -----NRP-----41
DB 721 ANIQMHVLDALGQLQPMGVAGELHIGGVARGVYLNQPLSABERFIADPFSNDPQARLYK 780
QY 42 -----W-----RGLVLSKIGKY-----53
DB 781 TGDVGRMLANGALEVLRGNDPQVKIRGRI-EIGEIEAALAKHPAVHEAVVTAREDPGD 839
QY 54 -----KLDQ-----57
DB 840 KRLVAYYTQSAEHTAVDLEALRSHLQOVLPEYMPVPIYVLEAMPLTNSGKLDKALPAP 899
QY 58 -----LKL-----61
DB 900 DGDALISRGYEAQCGEVETLLASINADVLKVEQVRHDFHFFELGHSLLAVKLIERMQV 959
QY 62 -----61
DB 960 GLSADVHLVFGQPTVAALAAAGGSEVVVPANLIPEHCDRITPQMLPLINLSQAQIDIV 1019
QY 62 -----61
DB 1020 VKDMPGVANVDIYPIAPLQAGILVHHISABQGDPTYLKALFALSADRAQLDDFSGALQG 1079
QY 62 -----ILRQLE-----67
DB 1080 VINRHDIRLATVIMWEDLPQVWLQRAELHYTELLDPDVGPDVEQLHQRDHRHRLD 1139
QY 68 -----67
DB 1140 VRTAPLRIVFSHDPVNDNRWLAMLLCHHLVSDATSLSVILOEQAHLLGQGNALGEAVPY 1199
QY 68 -----67
DB 1200 RNYVAQARLVSEAQAHEAFRDMGLGDIDEPTLPFGLQDVQDQGRNLEASVTLAEALNL 1259
QY 68 -----67
DB 1260 LRAQARQAGSVASLMHLAWARVLGNVSACEQVVFVTLLGRMQAGDADRALGMFINTL 1319
QY 68 -----TTIS-----71
DB 1320 PLRVDIGATTVAEGIKATHELTALLGHEHAPLVLAQRCGVAAPTFLFSALLNYRHSV 1379
QY 72 -----71
DB 1380 SDVRQKLPGLGSSLRHRRPYQLPVDLNVDDLGDFTLTIOAQVQISATRIGEMQVALR 1439
QY 72 -----TKYN-----75
DB 1440 NLVDALEHTPQAALNSILPDDERELLTGFNDTAHPYPRDVLIIHQLEQQAAQRPDAC 1499
QY 76 -----75
DB 1500 AVRGDSGTLTYAELNQANQLAHLRIELGVDPTRVAVSLRGAEMVVALGILKAGAYVPIDPDLP 1559
QY 76 -----VSKQPVK 82
: |||

Db	1560	YVPIDPDL	PDAQAYMLSDSAPRAVLTSHELLADLPDLGVPALVLDGRDDSA	LLKQPTG	1619		
Qy	83	N	-----	-----	LTMT 88		
Db	1620	NPAKALDLP	PHLAYVLYTSGSTGTPKGVNHEHLGVNRLWARDAYQVNSQDRVL	OKT	1679		
Qy	89	EP	-----	POYYILAGPIQNVSYTYLWF	-----	110	
Db	1680	PFGPDVSW	EFPLLTGAELVMARPSCHODPDY	-----	LAQVISDAGITLLHFVPSMLD	VFL	1737
Qy	111	----	DFYSTQLRK	-----	-----	-----	119
Db	1738	EHRSTRD	-----	POLRRVLC	SGEALPRALORRPEQHLKGVELHNLVGYGTEAAIDVTAMECRP	1795	
Qy	120	----	-----	-----	-----	-----	119
Db	1796	TDPGDSYPI	GRPIANIQWHVLDALGQLQPMGVAGELHIGGIVGARYLNQPOLSAERFIA	1855			
Qy	120	----	-----	-----	-----	-----	121
Db	1856	DPFSNDP	QARLYKTGDVRMLANGALEYLGRNDPQVKIRGLRIEIGEIAALAKHPAVHE	1915			
Qy	122	----	KVYISQYNHTAK	-----	-----	-----	133
Db	1916	AVVTAREDI	PGDKRLVAYYTTAEHTAVDLETLRGHLQVLPYMWPAIYVYLEAMPLTS	1975			
Qy	134	----	-----	-----	-----	-----	133
Db	1976	NGKLDKRAL	PAPDGDALISRGYEAPQGEISEQIATVIMQDILLVGEVGRDHDPFELGCHSL	2035			
Qy	134	----	-----	-----	-----	-----	133
Db	2036	LAVSLIDRL	RKQDLNLNVNTVFTAPS VREMA LAISQEQALFRAPANRIPAHCTQLTPDM	2095			
Qy	134	----	-----	-----	-----	-----	133
Db	2096	LPLVELSVAQ	IELTASAVPGAANIQHYYPLAPLQEGILFHYLLNLDERDAYLVRSTIEFD	2155			
Qy	134	----	-----	-----	-----	-----	139
Db	2156	SRARLDAF	LEGLQTVIDRHDVLRSSVHWVGLPQAVQVVRHQASPHTLTLRDEDEALSQ	2215			
Qy	140	----	-----	-----	-----	-----	139
Db	2216	LDRLSDP	RLRLDLRQAPLLAYIARDPDSERWLLALIDHHMISHVTLLELIEIRLWM	2275			
Qy	140	----	PP	-----	-----	-----	141
Db	2276	RGOSAE	LLPPQYREFVAQTLASPSAHEAYFTGRADVDSPTAPFELLEVOGDNDVBE	2335			
Qy	142	----	CGRVPSMT	-----	-----	-----	149
Db	2336	SELALSSDL	CARINTQARERGMSPVLFHVANAQVLARCTGRDDVVGTAVTRLOQTGLG	2395			
Qy	150	----	-----	-----	-----	-----	149
Db	2396	AERAMGMFM	NLTPVRVOLATQSVQELVMATHRDLSELLSHEQASLALARCQSSVATGVPL	2455			
Qy	150	----	CLIS	-----	-----	-----	152
Db	2456	FSSILNRYH	QNEDSQLQWPGRLLDGAERTNYPLCLISVNDYDGLGILLIHSVQPADPQRL	2515			
Qy	153	----	EMLVNSKRN	-----	-----	-----	161
Db	2516	CAMQCALEQ	LTALAHTPKEVTQDVLPAERNLILLETFNTRQDYPTDLCIHLFEA	2575			
Qy	162	----	-----	-----	-----	-----	161
Db	2576	QVRTQPD	AIAVAFOAQRLSYAELNRQANLAHLIGLIGCPDDRVAICVBERGVEMVGLL	2635			
Qy	162	----	-----	-----	-----	-----	161
Db	2636	GVLKAGAA	VYPLDPYAPERLAYMIEDSQPALMTORHLOEYLPITLPLVLDDDKORT	2695			

QY	162	-----DT-----	163
Db	2696	FTERDDNPVVEALGVRNLAYVITYTSGTGPNPKGVMIEHRRGLVNVSYDAARLFDLSPTDTV	2755
QY	164	-----	163
Db	2756	LQQTMLNFDLSVEBIFPALLAGATLTPSRBIFGSEGTENGINPTVLHLTAHHHTLVAE	2815
QY	164	-----	163
Db	2816	WHKQPOVABQRLQHVRLINVTGDALSQAQKLWDEVPAHTRLINTYGPTEATVCTAA	2875
QY	164	-----GEOCGNFTTFNPMFNPVRWNTKLYV-----	190
Db	2876	VSHDAAAGSESGNATIGKPM-----NTRIYLLDAHQPVPYGVAGEIYIGDGVARG	2929
QY	191	-----	190
Db	2930	YLNLEYNABERFLADPFSESPDARMYKTGDARYMADGRIEYLGRNDFQVKVRGFRIELG	2989
QY	191	-----	190
Db	2990	EIEARLGNCTGVKEAVVIABEDNPGDKRLVAYVVAQPOSQLTAAIDLAEALPQLAEMYLP	3049
QY	191	-----GPTKVV-----	197
Db	3050	SAFVLLDELPLTPNRKLDKALPAPDADALISRGYEAQGETETQIVAIWQDLLGIBQVG	3109
QY	198	-----	197
Db	3110	RHDHFPGLGGHSLAVLSIERLRKQGLNVLNVKTVFTAPTVMALAI SQKHVLFQVPAN	3169
QY	198	-----	197
Db	3170	RIPAHCTQLTPDMLPLVELSVAQIELITSAVAGGANIQDIYPLAPLQDGLFHYLLNRE	3229
QY	198	-----DSQ-----TIYPLGL-----	207
Db	3230	RDAYLMSMIETFSRARLDAPLEGLQTVIDRHDLRSVHMWIGLPQAVQVVHQAQLPVH	3289
QY	208	-----TALLRY-----	214
Db	3290	TLTLTPEDALSQDLRSDPGRLLDRLQAPLLLAYIARDPNRSERWLLALIDHMTSDHV	3349
QY	215	-----	214
Db	3350	TVELILBEIRLMRGOSADLLPPQPYRDFVQAOTLASPSSAHEYFTFRLADVDSPTAPPE	3409
QY	215	-----	214
Db	3410	LLEVOGDNDVEAKLALNSDL CIRINTQABERGMSPAVLPHVAVAQVMARCTGRDDVWF	3469
QY	215	-----	214
Db	3470	GTVVTGRLOCTAGABRAGMFMNTLPVRVQLITQGAQLVWATHRDLSELLSHEQASLAL	3529
QY	215	-----	214
Db	3530	AQRCSVATGVPLFSSLLNYRHQEDENRLQWPGMRLLDGTFRTNYPLCLISVNDYGSSELDL	3589
QY	215	-----AORN-----	218
Db	3590	IIHMQPANPQRLCAMMQCALEQTLDALAHTPQMAVTQLDVLPAERNLLETFNQTRQD	3649
QY	219	-----	218
Db	3650	YPTDLCIOHLFEAQVRTQPDIAVAVQGORLSYADLNRQANRLAHHLISLGI VPDDBVAI	3709
QY	219	-----	218
Db	3710	CVERGEMMITGLLVKAGAAVYVLPDPAPYABERLAYMITDSQPAALLTLPLGLQDRLPALS	3769

QY 219 ----- 218
Db 3770 MPLVLLDDEQYQGLAECDDNPVPTLGVRLNLAIVYITSGSTGNPKGVMIHRGLVNSVD 3829
QY 219 ----- 218
Db 3830 AARLFDLSQSDTVLQONTLNFDSVEEIPALLAGATLAPSREIFGSEGTGTHGIQPTVL 3889
QY 219 ----- 218
Db 3890 HLTTHWHTLVAEWNQPOAAEQRLOHVRLINVTGDALSQAOKLWDEVRAHTLLINTY 3949
QY 219 -----CTHSF-----
| | | | | YLVNA-
| | | | |
Db 3950 GPTEATVSTAAVSYSDAAAGSEGSNATIGKPMANTRIYLLDAHQPVVPVGVAGEIYIG 4009
QY 229 -----MSRNLFRVP-- 237
| | | | |
| | | | |
Db 4010 GDGVARGYLNEEVNAERFLADPPFSESPDARMYKTGLARYMADGRIEYLGRLNDFQVKR 4069
QY 238 ----- 237
Db 4070 GFRIELGEIARLGNCKGVKEAVIAREDNPEKRFVAYVVAOPQTOITAAELRBLAPQ 4129
QY 238 ----- 237
Db 4130 LAETWLPFAFVLLDELPLTPNRKFDKALPAPADDAFASREHVEPOGATEIALAQIWQSL 4189
QY 238 ----- 237
Db 4190 LDLERVGRHDQFPFELGCHSLLAMRLISQARHQLGVELGLAALFAHPEVSALAVATAQGR 4249
QY 238 ----- 237
Db 4250 SKPDIIVPVARDOAWPLSFQQRWLFLAOMEGASAAHYHYPAGLSLHGNLNLKALQALER 4309
QY 238 ----- 237
Db 4310 IVARHEGLRTFMQDGDQPVQRI SPADTGFNLMQHDLOGLADABEKLQALASESLQSF 4369
QY 238 ----- 237
Db 4370 DLOQGLIRGLIRMAEDHHVLLTLHHIVSDGWSVDVLTRELSALYAASFQDQDDPLAP 4429
QY 238 ---KYIN--- 241
| | | | |
| | | | |
Db 4430 LEQYLDYAVWQRWLSGVDLQOQSNFWQOTLADAPALLMPTDRARPALQYAGAALPV 4489
QY 242 -----CTKLK- 247
| | | | |
| | | | |
Db 4490 VFDKOLTRGLKALSQRSGTLFMTWMAAWAGLGRLAGQDDVVGITPVANRTRSEVGLV 4549
QY 248 ----- 247
Db 4550 GLFVNTLAIRDLSKPTAETLLARVKQOTLDAQAHQDLPPEQVVEVINPLRSLSHSIF 4609
QY 248 ----- 247
Db 4610 QAMLSWENNEASDLTGLDMTLKSIELAADTAQFDLTLDMAEVDQLVGTLEYATALPDES 4669
QY 248 TMRK-----LKRQAP-----VKEQFE 264
| | | | |
| | | | |
Db 4670 TMRYLGYFORVLEAMVADQOVLEHVALLGADREHLLAGLNATEAPFPQDRTIHOLFE 4729
QY 265 KKAK----- 268
| | | | |
| | | | |
Db 4730 ERVQAQPDIAVAFCAQPLSYAELNRQANRVAHHLISLGIKPDRVAICVERGVEMLIGV 4789
QY 269 -----KTQST 273
| | | | |
| | | | |
Db 4790 LGVLKAGAAVPLDPAYPAERLAYMIEDSTPSALLAQRDVQAHLPDLDLPLVLLDEQRT 4849
QY 274 T----- 274

Db 4850 TLSERDDNPVETLGAHNLAIVYITSGSTGVKGVMEHRGLFAVSAAEQYALHAPLN 4909
QY 275 ----- 274
Db 4910 HLQAGFFDVFSAADLIRSLAFGGLTLCPRETLMDDPPALYRLLSEESIGFADFLVALVNA 4969
QY 275 ----- 274
Db 4970 LLGWEEETGHDLSFMRTVVCSDIWTASARQLRKLCDGHVQVWQAVGVTEASIDSTCFE 5029
QY 275 ----- 274
Db 5030 FEATSQVDGVLPIGRALANTRIYLLDELQAVPFGVAGELYIGGAGIARGVNLPLQTAE 5089
QY 275 ----- 274
Db 5090 RMDNPFVAGERLYRSGDMARYRADGNI EFLGRNDSQAKLRLLELGEIARLAEVAGV 5149
QY 275 -----TP-----YF-----SYTTSA 284
| | | | |
| | | | |
Db 5150 RESLWIREDSGGTPKLIAYFEVATREDESGPALTPRALRQOLQLNLPYMI PAAFVRMA 5209
QY 285 ALNVTN----- 291
| | | | |
| | | | |
Db 5210 ALPLSANGKLDRRALPEPDADAFDQHDFAADGFELETAIAAIWADVLGVAQVGRHDDFFA 5269
QY 292 ----- 291
Db 5270 LGGHSLLVMRLVAQROQLNLEVSPVFPAAVLRQFAERLGNTODNARVAIKPVORSGA 5329
QY 292 ----- 291
Db 5330 LPLSYAQRLWFLAQLGEGSAAVHI PAGLRLGNLDQASLQALDRIVARHEALRTTFVQ 5389
QY 292 ----- 291
Db 5390 EQQPABEORISAAETGFRILQVLQAGOTDAEDKLLAIAOEAASEHFDLVNGLVRGLR 5449
QY 292 -----VTSI----- 296
| | | | |
| | | | |
Db 5450 MANDNHVLLVTHHILSDGWSGVNLRELGALYAPROGAGDPLPALPVQYDYALWORD 5509
QY 297 ----- 296
Db 5510 WLSGDVLQOQRQYWOQALAGAPALLTLPTDRPPAQDYSGQTLLEVLDTQLTRGLKALS 5569
QY 297 -----TTAARRVS-----TSTIAYRPDSS 315
| | | | |
| | | | |
Db 5570 ORHGSLEFMTVMGAWAALLGRLSGQDDVVGITPVANRMSAEVENLIGFFVNILAIRVDS 5629
QY 316 ----- 315
Db 5630 GTPSVEALVRQVKORTLAAQANQDLPFEQVVEVVQPQRSLSHSPIFOAMLSWQNNEDTAL 5689
QY 316 ----- 315
Db 5690 VLGDLTQVAVAGDTAKFDLALDIGEVDGQLIGTLEYATALPDESTWRRYRGYFLRLLE 5749
QY 316 -----FWKSI MAT----- 323
| | | | |
| | | | |
Db 5750 AMAADDOQVLEQVPLDPTAEREYLLKDNATERTYVPVQLMHRLFEAHAEAPQAIARVQ 5809
QY 324 ----- 323
Db 5810 GEQTLTYAELDSRANALAQHLRKHGCVGTRVAILLDRSVELLASMATLKCAGAYALD 5869
QY 324 ----- 323
Db 5870 FLAPEERLRFMEDESEAIMLLSRDLTAPDWTPLRLDITLLELSALNQEPVVPVATEVAGET 5929
QY 324 ----- 323

Db 5930 PACIYTSGSTGPKGVIVTHNGIVRLVQDNGYDFSAEDRVAFSSNPAFPASTPEIWGA 5989
Qy 324 ----- 323
Db 5990 LLNGCQSVIIEBPQVLEPVAFAALLKRHGVAMISSTALFNLYAGLIPALAGLRMWC 6049
Qy 324 -----QUR----- 326
Db 6050 GERADPASFRVRREHSAQVRLFNQYGPTEGTTCAYSBIFDVLPTLSLPICKFNANVRV 6109
Qy 327 -----DLA 329
Db 6110 YVLDARBPVPMGVVDIYIGGTGVALGYLNRPELTAERFSEDPFSQAGARLYRTGDLA 6169
Qy 330 TWV----- 332
Db 6170 RWLPDGNLEYLARNDQGVKRGFRVELGEIBSVLHLYDGVNRVSVVAHEASPGDRLVAY 6229
Qy 333 YT----- 334
Db 6230 YTVHAGVAAPDEDLRAQLSADLAEMYVPSVFVRDLALPLTLNGKVDQKALVPDLNAMF 6289
Qy 335 ----- 334
Db 6290 ERTVAPEGATEQALAEIPOBELLGLERVRHGDGPFELGCHSLLAQLVSRVRQQLNGDMA 6349
Qy 335 ----- 334
Db 6350 LRQLFNHPTVAELAKVVDGLQTVDSDSIEPIERNAPLALSFQQRLWFLDRLDPGASSAY 6409
Qy 335 -----TLR----- 337
Db 6410 HPMWSLLRGELDHRAKALDLVARHESLRTTFELHGEQPVQVIAAADSGFALAEDDL 6469
Qy 338 -----YRQNP----- 342
Db 6470 RSQYEQASLNASRIADSEAAAPFDLRQGLIRGLRLRLADDEHMLLITQHHIISDGWSV 6529
Qy 343 ----- 342
Db 6530 GVLINEFTALQATEQRPDPLPALSIOYADYAAWQRTFTGERLAEQADLWREHLGGAP 6589
Qy 343 -----FCB----- 345
Db 6590 TLLSLPTDRPRVQSVYRGGAVPVTIDAALHORLERFCQAHNVTLFMGLLSAWSVLMTRL 6649
Qy 346 -----PSRN----- 349
Db 6650 GNERDVIQVPSANRGRTETENLIGFVFNALALRVDLTQNPSPAQLLEQVRQTTLAAHEH 6709
Qy 350 ----- 349
Db 6710 QDIPFEQVIEALQPPRSMSHSPQLCOVALSLDNTSIGGELKLPGLSLHPVLOAHETAQFDL 6769
Qy 350 -----RTAVSEFMKNTHVLI----- 364
Db 6770 MLTLASENGALNGVIEASDLFDRSTVERFAQHFTLLEAMVEDVAQVGLGLPLLSPAQR 6829
Qy 365 ----- 364
Db 6830 LASPALLQPKAVPASGLMVHQRFEQFAAAHPQNTIALVFGREHVSQALNRDANRLAHELL 6889
Qy 365 ----- 364
Db 6890 AQGVPRDPRVAILAERGTMICAVLAVLKSAAVYPLDPTPTERLGLVLLTDSAPVALLA 6949
Qy 365 -----RNETP-----YTIYGT----- 375
Db 6950 QSACLDVLPASVPLTLGADNDSADLADESVDNRNPREPAALGLSREHLYAVIYTSGSNGL 7009
Qy 376 ----- 375
Db 7010 PKGVLEHGNVARLFDATAGFNFCHEDVWTFPHFSFAPDFSFWELWGCALCYGKGLVIPS 7069

Qy 376 -----LDMSSL----- 381
Db 7070 EVARSPDDFVALVCEQQVTVLNOTPSAPROFIQARERSPOSHALREVVGSEALDFRSIQ 7129
Qy 382 ----- 381
Db 7130 PWTARTPLSRTRLNVMGITEITVHATYVPIQSEIDTAMPSLIGPALDDLCRLRDDYQ 7189
Qy 382 -----YNNET----- 386
Db 7190 QPVPVGNGEIYIGGAGVARHYLNRTELNAERFIADPYALQSGARLYRTGDVAHYRSOGG 7249
Qy 387 ----- 386
Db 7250 VVNVGRNDSQIKIRGFRIELGEIEAQLLACPEVREAMVILREDRPGDKRLVAYLIAEDGA 7309
Qy 387 ----- 386
Db 7310 APESALLRSQASVLABHMLPSPFVTLTETWPLTTNGKLDRAALPAPDQFATVSRDYEAPL 7369
Qy 387 ----- 386
Db 7370 GAIEITLAAWQELLGVERVGRQDHFPELGCHSFIVISLIERLQIGLGLLDVSTVFSAPT 7429
Qy 387 ----- 386
Db 7430 LQAMAAVLGAGTAABRVPANLIPVDCALTDPDMLPLVKLTRQBLEHIVADVPGGVANVQ 7489
Qy 387 ----- 386
Db 7490 DIYPLSSLOEGILPHLLQSEGDAYLMRTLATFDSRALLDKFALQVVINRHDMRSSL 7549
Qy 387 ----- 386
Db 7550 RWQGLPQPVVVRQAOQLPVIQDTPAGDALQMLRERTNTHYHRLDLQOAPLIAAYITY 7609
Qy 387 ----- 386
Db 7610 DTRQEKWLMALLDHLISDNVTLRLIMGEIQAVMDGRADALPPSQPYRNFITARAACVSA 7669
Qy 387 ----- 386
Db 7670 EHEAYFRQLLGDVDTTTPYQVLDVRGDDATILRSVQDLSDDLARIHSTARAQVPTSV 7729
Qy 387 -----MFVE----- 390
Db 7730 LFHAMGLVVAATSGRDRDDGIFGTVLSGRSQCTSGANHALGMFINTLPMRIRLQONSVRDI 7789
Qy 391 ----- 390
Db 7790 VODAYQQLSGLLTHERAPLALQRCASAVDASLPMFTVLNCRHGDVNTAGENIEDMGEE 7849
Qy 391 ----- 390
Db 7850 QEGVHFLGSETRTNYPIEIAVANEANGFSLTAQSIDGIDPHRIAAYLQGAELVNALEQ 7909
Qy 391 -----NKTASD----- 396
Db 7910 DPARLASSLEVIPOAERQQLLNDNFDTASDPAPVPIHALFEDQVRNRPDAVALVYEDRQ 7969
Qy 397 ----- 396
Db 7970 LSYRQLNRANHVARQLQLGVQPDERVAICABERSLDMIAGLLGVLKSGAAVVPIDPAHP 8029
Qy 397 ----- 396
Db 8030 ADMRAFMLQDSQPRALLTOGALSIPVGDTPMLMLDLSAESLLAADQAFDANPVVDGLTAE 8089
Qy 397 -----SNKTT----- 401
Db 8090 NLAYVIYTSGSTGQSGKGMVMEHRSVFNFVNLTRTTHOHCPTPATVALNAGFFPDMSIGK 8149

402 ----- 401
Db 8150 ISQLFSGHKLVIIPIQLLRANGSELDFLEAHQVHAFDTPSQLODTLLSAGLLERSSQVP 8209
402 -----PT-----SPSMG----- 408
Db 8210 SVLLGGEAINASTWEKRNCPTRLNYMGPTCTVDATIDLRLDGEKSGIGPIANVQ 8269
409 -----FORTFIDP----- 416
Db 8270 VHVLDARGEPAPLVAGBIHIGGSGVARGYLNRLDELSAERFIVDPFSDAANARLYKTGDL 8329
417 ----- 416
Db 8330 GRWLADGTLEYMGRNDFQVKVGRFRIELGETENVLLAVPGIREVVVIRNDSQGDSDSQ 8389
417 ----- 416
Db 8390 LVAVYCGESVAAEHLRSELLRHLPEYVPSAFVOLDALPLTANGKLDRRALPAPQDALA 8449
417 ----- 416
Db 8450 SKVYEAPOGTEEAIAEIKWALLHLDQVGRNDGFLGSGHLLAVOLLRLRKLGTTRIT 8509
417 -----LW--DYLD----- 422
Db 8510 LRELFDAPTVRGLASLVNAAAPGEAQSIPRANRSGRLPLFSQORLWFLDHLDAAGAY 8569
423 ----- 422
Db 8570 HLPWALRLTGALDTAALEATLDRLVARHETLRTFELVDGEPVQKIAPADSRLLPKQL 8629
423 ----- 422
Db 8630 RDLGDBRTSTLARLQENATQLDLTGPMRLGRHLRVADAEHVLLIILHHIVSDGWSN 8689
423 ----- 422
Db 8690 SVLAQEVSAIYAFSOGQKDPLPALPQYVDVAAWQOSLDGPALQAQIDFWRKLEGAP 8749
423 -----SL----- 425
Db 8750 SVLNPLDLRPRPAIQSYTGVMVHVPSPALSADLRAFSQAQGSTFPVLLAGWSMLSHL 8809
426 -----FLDEIRNFSL----- 435
Db 8810 SEQTDVVVGTVPANROHPELPMIGFANTLALRVATDRETRLNDMLDRIKSLTLAAYNH 8869
436 -----RSPTY-----VNLTPP-----EHRRAVNLS 455
Db 8870 QDLPPQVVSALQPTENVSHSPLFQVMLSNDTPPSLLQLPDLVELLDSEHHTTQFDLS 8929
456 TS----- 457
Db 8930 LSLVDSSESGGLQYSSDLFDKTTAQAIMOLFARVLENMVSARQSIGQVLNDTPALPR 8989
458 -----NSLM----- 461
Db 8990 SAHTATIAATVEDEQPEALPYEAPEGETEIALANLWELLKLEKVSRRHDDPFKLGISLM 9049
462 ----- 461
Db 9050 AVQMASRLRKVLGKPIAVRDLFIEPTIAGPARTLDGQARPQHSNLVVRTGTSQRPFLF 9109
462 ----- 461
Db 9110 VHPLGGEVOYVRDLAAIDPOVPLYLAASGLAAGETPELYEVPAMAAARYLAIRQVPKG 9169
462 -----W-----WL 464
Db 9170 PYRIGGWSAGGLIAYEMARQLQSSGEKLEFTGIIDTSARLEQQAPALSGQFLMDWL 9227

RESULT 8

T15789
hypothetical protein C41A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15789
R:Bentley, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C41A3.
A:Reference number: Z18404
A:Accession: T15789
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7829 <BEN>
A:Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831.
C:Genetics:
A:Gene: CESP:C41A3.1
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/1;
/2; 7514/3; 7536/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:6547-6616/Domain: acyl carrier protein homology <ACP>
F:2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 20.3%; Score 498; DB 2; Length 7829;
Best Local Similarity 3.0%; Pred. NO. 0.59;
Matches 223; Conservative 92; Mismatches 140; Indels 6916; Gaps 74;

QY 1 MG----- 2
||
Db 366 MGFVNEEMELNRAISSYFGGTNACAIIEKPKPSLVQKESYAESNVLFSAKSHSLK 425
QY 3 ----- 2
Db 426 LQIEEYTFQMAQSDSAMEDIYTVNERKTYDFRAAVFGKDNBEIARKLQDGYSLTNLQ 485
QY 3 ----- 2
Db 486 ESTFEVEFGSGNEKMLLRLMLEKNETFHTSTVDYKYLKAEKTCGFPPEARFALFPFKLTLT 545
QY 3 ----- 2
Db 546 PLTYNVSRLISSMATPELLVQVNTLPNKLGRGLGQIFCLAVAKVITPESAVQLIKGVVA 605
QY 3 ----- 2
Db 606 EANLTDLGDIELKSSKIPIEIOHLKSTKKKILPIHISGELKETAKPNLMTFIVNGEIL 665
QY 3 -----RK----- 4
||
Db 666 ELDPVRKVKQLICQLPACGDFPAVKFRGRIVKTPYVFLKQFWEVQVTAMTNIVDEQT 725
QY 5 -----EMVVDVPMF----- 15
||
Db 726 NSLSDAEIESTVRITVQKFLDIEEDDINLLETGAVDLSITSEMVEAFGTAVNQTMPFDL 785
QY 16 -----VLISISFL-----VSEINC----- 30
||
Db 786 LEAYPTILNIVDFLTLVTVPTVKATTSIHKKTSSELSTSDINVIACDYQFAGVEGEKEL 845
QY 31 -----K 31
Db 846 WDTLTSRLTTGKISDIRKQCEGDAGLEVGLLKQDISMFONSFFAIKDAEAFDPOHR 905
QY 32 VMKALYN-----RWRLV----- 46
||
Db 906 LLLNAYNALEKSGLTSPDADFLAISAHSEYRALAEKHINELDERLWMGTVHSMVAGR 965
QY 47 ----- 46
Db 966 LAVLMGIRGRAMIVDTTCSSVATALEMAVKSIREGRKFAIVATSQLIQSSKWLKSLTLL 1025
QY 47 ----- 46

Db 1026 DHHSTNPSVDGSGFCRSGVGVILLKTAEGDSAVIKISSAKSHHCGAVMTPVVSISQ 1085
Qy 47 -----L 47
Db 1086 LLEEAGSPSYVEGHGTATSGADSAESMAYQKLGSELINSSVKAQFGHCEVASGLIQLMKV 1145
Qy 48 SKIKGY-----K 54
Db 1146 SSIKGHGIIPSVHNILPSEHNRNENIRLPFAEEKIDRSIAVSGFITGCTKTVTYTER 1205
Qy 55 LDQLK-----BILROLETT-----69
Db 1206 VSQLNDNIQNCYLLPVSAKTKDGLKACLSLIEHMDNSCESLYDISTLQOKQKNPKWR 1265
Qy 70 -----ISTKYN-----75
Db 1366 TAVGSSHADVVLKQLTSEHNTSLTNWHISTSHHSIGCSTPFHNIPEPEDHYSMFCH 1325
Qy 76 -----75
Db 1326 RLRPEPHSNTNSIYHLLAVVVALIRVILKHKLNTNSFAVGFSNLIIVLAADAPSHYLN 1385
Qy 76 -----VSKOPVKULTWNT-----89
Db 1386 LLHAFANDDVMMKRIARDVTIISIVKLLNLNGEPITTARQAVEATIDQKVVRILPET 1445
Qy 90 -----PQOYIILAG-----98
Db 1446 TLILSPSRAYEFASLETIQDYKLGKFTQGGQVDFAGTGPILKIDLPEYFPNRKS 1505
Qy 99 -----98
Db 1506 FWLPIDPSVPSNEKEKPLIPKSYEFLLKSQKHQVHVVDKIVLPGATSIRLVHQLN 1565
Qy 99 -----PIQNY-----104
Db 1566 GKPTVELSNIDFLNKITPSEAPSVVKBEEQGLEKLVGETDAISFKTELQNFNPINE 1625
Qy 105 -----ITY-----107
Db 1626 RLNAEVHHTDNIYERFANSHLTYNREFQWDSLKYTMCKGVRVFSMKDLIDLIDGTLOA 1685
Qy 108 -----107
Db 1686 IVGCYFENTDNSPFVFTIDQLSILNGDISQQLHAVLKYDSSGNFINGDATVYDALG 1745
Qy 108 -----LMFD-----111
Db 1746 NIILHISNVTFKRLNGQAPSLTSKTVDSKITYKKVENEQKRASKMHLHVWFENFGWT 1805
Qy 112 -----PYS-----114
Db 1806 DIDNTGFDLGLTSIQAVKLRNAIKSNYPNASSTCVDPYSIDLGLSTLNDPQVTE 1865
Qy 115 -----TQLRKPAC-----122
Db 1866 TSTGEDDIQDLTDEHKPTRLAENPIGWMAAACRLPGGVSPSELWELLKIGKNASSRIP 1925
Qy 123 -----122
Db 1926 ATRVPTRTLISGSKYGNPVEGNFITQDVTQDPDPFKISBAELIDPOORLLLECVQ 1985
Qy 123 -----YVYSQY-----128
Db 1986 ECLNSGVITSNVGVFVGLMEKEYQDMMESSILAMLGSAVAIAGRNVYIFCYGPSV 2045
Qy 129 -----128
Db 2046 TIDTACSSSLVALEMAINALLDNCRSKVIVAGVNLILNEKQGLRTNGKMLSQHGMSLSP 2105
Qy 129 -----NHTAKTITFRPP-----140
Db 2106 DSRASGYGRSGCVVLMLELAKPNFHYMSTTQSVNVNHHGRSVSLTAPNGVAHROMLTTSV 2165

Qy 141 -----PCG-----143
Db 2166 INQSPSLAIDYWEAHGTGTPLGDPIDFENTLSSILQNIIGSVKASLGHEASAGTCGLLK 2225
Qy 144 -----RVP-----146
Db 2226 LFLMLTYQYVPTLIHFHVVLNKDINAGSIRLPIIGEDSELVSAGISSFCVSGTNAIAAFN 2285
Qy 147 -----146
Db 2286 DNNKLEPIPIHKYIILPISAKNQISLDNLEKQILSVIPLTDVPICNIASALANNRSHFT 2345
Qy 147 -----146
Db 2346 IRNALIVSNGIVNSKMEGKPHRVAKKDRVHVVKLDCSLDASLLQYDVINETYTVASLKN 2405
Qy 147 -----SMTCLSEML-----155
Db 2406 PQSPAMKPAIKFUTSLSEYIEIVASDGEELLAVLLANGSLUKWENFNKMTIELPIGSLLT 2465
Qy 156 -----155
Db 2466 EFADHDLNSTTSSSIKSYQTPESHNLDSPELMKLMKLYITGYVDVDMATVYSPVEQFI 2525
Qy 156 -----155
Db 2526 ALPNYQFNKOTLWPEERLEIVDHYLIGTIDEESEDTLILKNQISLARHPQFKGKPLDVG 2585
Qy 156 -----NYSKRND-----162
Db 2586 TMSIATBALKIRNEIPFSIQNLKTELITLTKPAWLETNVTNRNEDDEGFMVSAIDGQRL 2645
Qy 163 -----162
Db 2646 FSLNASSVEIQNIEVPAVEVQIPDKVYVLYKECPNAVIRRHNNMYVDSRAEQSPRTANI 2705
Qy 163 -----162
Db 2706 VLNEIIGFAPTSDMFIEILGVLPSVHYMVQVDDGALWQFQMISQDKRVLSNIYVLKQAK 2765
Qy 163 -----TG-164
Db 2766 GLEIPTIRMHKSTLLSSQBASIVAAKTLOAVRHVKVCLAVGDVIBSGLDIDESQLSTGF 2825
Qy 165 -----EQG-----167
Db 2826 SELGIDSLATVDLLRLNQKYPPEIELTSDLPDNPSSIIDLIMIEQLLNEKGITEPSEP 2885
Qy 168 -----CGN-----170
Db 2886 NTPKTSLRGRKLSIPAVRAQVLAQIEFVENYNSKQEVQAEAPSSSECSNHLESDATV 2945
Qy 171 -----PT-----172
Db 2946 DRTEIRKVSIAVFDLATETLSAEDLQSKGFTLGMDSLSIVDFVNLNDKYFPDDEITA 3005
Qy 173 -----172
Db 3006 SDIFDYPTVDELSDHIVRKXSSSVPPAASEIMKETMNGISTSDAETHKLENLSQSFMLL 3065
Qy 173 -----TFNP-----176
Db 3066 ENQNSINPTLKIWISNOTIKLVKPSDGNFLFELNANGQKEIKQKFTGPNNIIDLKGF 3125
Qy 177 -----176
Db 3126 HEGSTETLYMSLLNLVKSISKLEIQCRFGVQBFLGNSISRAFPMKTVAEEKNPLISPAW 3185
Qy 177 -----176
Db 3186 YQNQQVSVFVDSOSPITGNWLTGGLSGIGLEIGKFIANNGAENVILISRRQPTAKALRD 3245

Qy 177 -----MFEN--- 180
Db 3246 ILSTELTHIGLARKTIVLKIKNISAKLIVFQSKLSFSTKVKYISDLFSKXVTFYFNLIQ 3305
Qy 181 -----VPRWN-----TKLYVG----- 191
Db 3306 SKCFSKHPHWFLEWLEVYLIKQWTVHTIAADINDKEKIRELTKLANVGITGIHSAGVLK 3365
Qy 192 ----- 191
Db 3366 DSKIERQKESFNQVFTPKANGPHVLEIEKHFNKYKIENFIMMSSPTAACGNEGQNLNGV 3425
Qy 192 ----- 191
Db 3426 SNAYLEYQVQRRRQKSGCAIQGNWIDTGMATDENVRKFLANLGLFLOHNDKALKYLR 3485
Qy 192 ----- 191
Db 3486 ACILTKPELIMVANIDWNVILKNRDLPKDLINTGILPFEDFTGKMNESEPLSNGDFEK 3545
Qy 192 -----PTK-----VNVDSTI----- 202
Db 3546 VSMNFSVEDEEVELELLEKVVSSILMCSPTKLKNKNIMDMGLDSKLIVEFLNFINSTFK 3605
Qy 203 ----- 202
Db 3606 ISVNLSDAYNHPTLEKLAHIPEQMTIVDHPVNSVKSEBIPKSTDPFPIGINIFFDNKN 3665
Qy 203 ----- 202
Db 3666 DFDKAKSTAVKLENGEQLPTAGKYAVSVVGKSIKRDVWSKIKBAAPQIKLCQESSKCVL 3725
Qy 203 ----- 202
Db 3726 MLTGQSGQPMGRQLVENYEIFRTTLOSLCKKCDLQGVLSLWEILFNTDHYKLLQT 3785
Qy 203 -----YFL----- 205
Db 3786 KHMQPMFCGYATAQLMSLGIVPDYDLGHSGVGLVAGLAGIMSIEDGLRLIVERGA 3845
Qy 206 -----GLTALL----- 211
Db 3846 MENIAGLGNLAVOREIADEVLRKPKSVATINSPKQVVPAGTSVLDAALAFVKGQKQ 3905
Qy 212 ----- 211
Db 3906 ATYVNOQYPPHSHLIQETHLVSLRQCLADIKFSAGRTPLVSNVTGIIINTPSEAVIVKHT 3965
Qy 212 ----- 211
Db 3966 VSAVKFVDCVETLOAKGVTVWIDAGSAVALATFVKRIIOPTELSKHRIVQCKEKESDVD 4025
Qy 212 ----- 211
Db 4026 NLVQACLEQSGLPISWTLVYGCGRNADERLVEFPMTHNDIIKNDSEFELLEGHQNGKI 4085
Qy 212 ----- 211
Db 4086 VVAGAYQLFKTDQLVKLKAAGMELMLKNVKFLKPWYIEDNRQYQIOWNSDMTIELIVNSV 4145
Qy 212 -----LRY-----AORN----- 218
Db 4146 IVCSLEVEPQNSVLKLETISENEKPFVHDHYETLFRNGLOYDSGFRRIESARRSKRCF 4205
Qy 219 ----- 218
Db 4206 SQIKSPFAPWPLIDSAMHSITASVPPRRPCYFLPVGMSVTMKDNTNSFTLPLNHAQTVI 4265
Qy 219 ----- 218
Db 4266 TSETDKFIQNVALLAGDTPICEVRNMTIIVLKTPEVHTRIPNSIETVETPPKSEIEVG 4325
Qy 219 ----- 218

Db 4326 FDISLPYNOISENSEWQHLLKNTNTVKQKLHNRSCLKQDHARVALLDSLDARYWDPEYFGIRP 4385
Qy 219 ----- 218
Db 4386 SEAKFIDPQORLLLCVAKLLDSLLITSLTNTGVFICGSANEFSHIVAYGYKDPRAEW 4445
Qy 219 -----CTHSFYL----- 225
Db 4446 SGGTSNSALAGRIAHWLKGPVVTLDTACSSSFYALSAACDALRTGCEYAIIVGTWNLV 4505
Qy 226 ----- 225
Db 4506 MHEMTVDVLQNAKMTVDDFCFAFDVANGYKRSSEAVCSMLITKSPNIDSVATITNYATGH 4565
Qy 226 -----VNAMS----- 230
Db 4566 NGTSSSLFTPNGLSQLEVQMPATNPLEKILEIQTHCTGTGKLDPIEINAKSLVSSACKI 4625
Qy 231 -----RNLFRV----- 236
Db 4626 GSVKSNIGHTEGSGLVSLCSSLMSFRSKYRVAQLHLKCPNTSIKTNKMICRPFIGEDADE 4685
Qy 237 -----PKYI----- 240
Db 4686 NNSILINNFGTGSNCVSVLKPNAISEHFVSSEVFYPILLSSHSAKSLQKYVQVLCEFI 4745
Qy 241 ----- 240
Db 4746 SNSAKSLHDIMMSLFQKKIHVHRQFIIFNPKRIAVTSLDGFVEVRDERLEKLEKHPCSVF 4805
Qy 241 ----- 240
Db 4806 LKSGVHFDKDKFORVDLPISIVFNNTLHWALDSYRDEIDRHESQOMSPKNIFFEYKVLLE 4865
Qy 241 ----- 240
Db 4866 TPMPQQNPISKVVCIGRLDLIPKIEIDEVSSPANGIIIVFHPISNSIFEYLKLAQVWSLIS 4925
Qy 241 -----NGTK----- 244
Db 4926 RNQNVFIIICCFENGTSHTWTGTLASLASEKMIPIYKPVSIDKVDALKLEFNHEYMFEAI 4985
Qy 245 ----- 244
Db 4986 FYSKRGYVERLAKVNPFPKAPQBYEKCLISGGTGGIGSAILNELPKPKSSVITRKNIAS 5045
Qy 245 -----LKNTMR----- 250
Db 5046 EDGKTPLSDDITRLDISHKFNYYVHLAGIVNNSLHENVKRDSDIDEMVSIKLOGAKNLMKC 5105
Qy 251 ----- 250
Db 5106 CDETSHFVSSSIANVLGSGQSNYAFNSGLVTSFLETSTKSTIIHGWPKWGVGMLAQP 5165
Qy 251 ----- 250
Db 5166 ERREIVKQIESNGWKLPPNQDAISVFTYQPMETHEQIIVFDGDFDTIVARQPHLOKLSE 5225
Qy 251 -----KLKRX----- 255
Db 5226 VVEKTKVKEEIKKSLNFEIFEIVGITDISKLNIPFMDLIGDSLCLMENLRYSLKNKF 5285
Qy 256 --QAPVKEQFE-----KKAKTKOSTI----- 274
Db 5286 DLELTVSEMFENATYQKLTQYVETURKAKHSESLSHDRVSSQVSNKEDDTRVAVIGMSAE 5345
Qy 275 ----- 274
Db 5346 FSGSSNIHEYWENLMDGICSTGNKNKYLKPNPGFDNKFNLTDDEARVLDQVRFKFIQHA 5405
Qy 275 -----TPYFSY--- 280

Db 5406 YLALENSGYVQKHLCVGFAGAPSDYGRADDDHDDAMRKLFVNMNMSYLASYCLD 5465
Qy 281 -----TTSALNVTNTVYSITAA----- 300
Db 5466 LKGEAVSVYACSTALVAVANAVKSIQSGMDYALVGAASIAEVSGALSGDFDQKKTWFS 5525
Qy 301 -----SEWKN----- 359
Db 5526 KSGVCRPPDKDSBEGIVRSGVGCFLKRYSOALLDNDNVHVIKDFAINNDGHSRSPMA 5585
Qy 301 -----RRVS----- 304
Db 5586 PNPAGOLKCMTDVLARPTNKERISFVECHATGTTIGDTIEMNSLRTAYSPKNAIGS 5645
Qy 305 -----TSTI----- 308
Db 5646 CKANIGHAYASGLAALVKCAKMLQTGIIPQVNPSEFRDGMGQFFTVNGKKSITSONSL 5705
Qy 309 ----- 308
Db 5706 ISIDSGIGGTNVHNVIEFPARSOBWKISSENILYDMIPISAKTEYSLDHTSEAIKY 5765
Qy 309 ----- 308
Db 5766 LQDTNKIAQCSSTFIHSRVPMDSRTYLSVNNNELLKIRTNKKTWFKGSPKIALFPAP 5825
Qy 309 -----AYR-----PDSPFKSIMATOL--- 325
Db 5826 QGIOFTNLPNEYLKNSAYREVEYLCBLASSFGIPSLGILYPTKQNFDLIHATQPAQI 5885
Qy 326 ----- 325
Db 5886 AIFVQCMAIPKAIRKNVFNPTCLIGHVGEYAAAVISGALKTEBALKLIKESLIGKTEK 5945
Qy 326 -RDLATWY-----TTLRVQ----- 340
Db 5946 ARMLAVWNYEKQLPSHVHVSIIIDANTKCVVGPVETIDNLEKYFINNHKIRYNIETKHGP 6005
Qy 341 -----NPFCE----- 345
Db 6006 HSKMFHCISKEFEFCESFATKVPLIPMISITGSEIKIPDSKYCTMHLTPNVNLELVVD 6065
Qy 346 ----- 345
Db 6066 HIMKLDIDIIIEVVGPTGVLNLLAKRNSKIVVWPTCGTKHKPKISLGBECIQGLWSNGVDI 6125
Qy 346 ----- 345
Db 6126 RKLTPKLAVDQVPGYCFDERQFGEKNQKISQKLNRFYKESKWPVSSELEETKPVRFPT 6185
Qy 346 ----- 345
Db 6186 VCDGSLNGISEPLVYLLARKSEDIHKNYPDVELFLKSSPRCSTVVFGMDNSPSVHLSLG 6245
Qy 346 -----PSR- 348
Db 6246 LIRCQVLSRIDLYKVENFENIAMSVIQVLTANGLYFRIDSTGLCKGHPRTTEIPSRL 6305
Qy 349 ----- 348
Db 6306 HPVSRGHALVFGANGFICGIVFRLLQEMGMNVIPIRASIPSCDITNIKOVNVFKSLGP 6365
Qy 349 -----NRTAV----- 353
Db 6366 KKPSVINCVGVEVSANAKNTSLEQEIVLSPTFGSVNILKLEBFSIEVDKLVNPSLS 6425
Qy 354 ----- 353
Db 6426 SWPLLGNDYASANCFVEALTKQSKYIKQFLTLSLPLEGRMYESSSTSTKHNLQI 6485
Qy 354 ----- 353
Db 6486 LMQOELTNVLKTAPESNMNGIVFVSSVNPADIATKSIAPHRTDENGSEVKAVESLKLPK 6545

Qy 354 ----- 353
Db 6546 STSCFVIABWKETLGLSILNDANPNPFFSLGGDSLSALQVVMKVQKKTDRIVDVNDLFD 6605
Qy 354 -----SEWKN----- 359
Db 6606 NPTLOBFTKPVKNLTTTKFPAGNTWKISYDAIPLTNSQTMFMLRQIDTTSKYNLIFKIT 6665
Qy 360 -----THVLI-----RNETPTTYGTL----- 376
Db 6666 ISYETKFWBFLKYLHSLIAYQPSYRTVPKSGNSPYQICSLTESPHDPKRCNLNNAI 6725
Qy 377 -----DMSSLY----- 382
Db 6726 SHEPNHLFEICKSTPLRVRVAEDCDNRHIIHFNQHHILTDGWSMTVLSDTVSSLYAYR 6785
Qy 383 ----- 382
Db 6786 GETSPPSKTQTIQSIVAMGTKSSGDIKEALEYYQNTVHTIIPYDSEGTGNTSPSYVRISKL 6845
Qy 383 -----YNETMF-----VENKTASDSN----- 398
Db 6846 IPSKIWQKLVGLSKLYNTTWYNLALSVPDCAVRSTGQADILLAVAISGRNADNSELIGY 6905
Qy 399 -----KT----- 400
Db 6906 FMNALFKTSLPFEILRELEILNIVLSLEKRSFATIPFYQMVQONRKLNEISLFFNFR 6965
Qy 401 -----TPTS----- 404
Db 6966 QKLDYPTVMFGAKCEIEHLSLNNAFDSFTIDETPSLITVDKSKYLDTTVHMPAN 7025
Qy 405 -----PSMGFQ-----RTF----- 413
Db 7026 IFLKLANLRNMNTTIPIRTDPSLTFQKGLFTSWRLFTPEALLLSTNTVSYSDLAEKI 7085
Qy 414 -----IDPLW----- 418
Db 7086 ENISKDIQKQIAKATSVREDELVLGDLCKNSYFALLACVFLGDPYAPIDPTWPEPROLF 7145
Qy 419 ----- 418
Db 7146 VKSKVSFTLENCFCSCNLKLRNFNSRTQFSIYFTSGSTGVPKGVLMABEQSVSFMSTA 7205
Qy 419 ----- 418
Db 7206 SKQCMFRSNIRVLDSVKQVFDVSVNSIIGSVLNGVLISSHSHTITDLOKQCOYAPLPA 7265
Qy 419 ----- 418
Db 7266 AVFNGFTDKTMSRLESJETLTIGGETVSDVVIETAMKKFPRLRRTIYVPTTCIWSLTN 7325
Qy 419 ----- 418
Db 7326 KCKVSTLNGISALGDSLSNETCTICNNSVRGNVQVGISLARGVITSAPHGTPPSDIYST 7385
Qy 419 -DYLDLSL----- 425
Db 7386 GDIVDSKLSLOYIGRMDSQVKCKGVRINISIEKEILLCGLLQIUVVLYSNQMLIAFIV 7445
Qy 426 ----- 425
Db 7446 DQSKLLHDSLAVTKLNRTQIPDYFVQINKMPLNSGKDKSLLLQAPENIRKSYKREIV 7505
Qy 426 ----- 425
Db 7506 VMKNSLEEKVINVPSKILGRNVAPTDKPFESIGGNSLNAIQIAHRLABELKIEIKAHEILQ 7565
Qy 426 -----FLDEIRN----- 432
Db 7566 SNSLKTFCNTLKNSVQKPIKPNVNTKLRVPSNKNYILVHAIGGTIYPYPSFLQIFP 7625

QY 433 -----FSLRSP----- 438
Db 7626 KDLSLYGIEFDLKYFSNDLRELAHFAEIAAHAGNKKIFVNGHSMGGIMSRREIVABELKI 7685
QY 439 -----TYV--NLTPPEHR--RAVNLS 455
Db 7686 WGYDIPFVWLFDSWVLRNTELDIENIKQFITVYFSGLPDSEHRINRAIKLA 7736

RESULT 9
Ti6580
hypothetical protein K07E12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: Ti6580
R:Fulton, L.
submitted to the EMBL Data Library, May 1994
A:Description: The sequence of C. elegans cosmid K07E12.
A:Reference number: Z18540
A:Accession: Ti6580
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-13055 <FUL>
A:Cross-references: UNIPROT:Q09165; EMBL:U00054; PID:g485141; PIDN:AAA50715
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:K07E12.1
A:Introns: 46/3; 85/2; 201/1; 278/2; 470/2; 817/1; 927/1; 960/3; 1265/1; 1322/1; 1478/3;
; 6014/3; 6159/3; 6665/3; 7266/3; 7895/3; 8669/3; 8726/3; 9803/3; 10937/3; 12234/2; 1229

Query Match 19.5%; Score 479; DB 2; Length 13055;
Best Local Similarity 1.7%; Pred. No. 14;
Matches 216; Conservative 86; Mismatches 135; Indels 12521; Gaps 71;
QY 12 PMFVLISLFLVSP----- 27
Db 113 PKNATSPKSLFVKVHCDYGVEMDEVDTMFCQDKKWMTTPMCRGGQGLCAADNGGCSH 172
QY 28 -----INCK----- 31
Db 173 TCISYNDEKIECKPRGWTLDVEKTCIKPKSLCSLSGCTCNGITETQFACSCGNK 232
QY 32 ----- 31
Db 233 QKCLLIAGPRRIYIEPGPYEAPGPNINISCTSVAYFPDIYFKNQKVNTPDQDQNTL 292
QY 32 -----VMSKALY----- 38
Db 293 RASQILIKBIYRNEEFTCVSDNIHGSANRTVSIWVTPGSGAPHLKASAGRTSLTYRWE 352
QY 39 -----NRP-----WGLVL----- 47
Db 353 PPSIINRPIITYTLTYNNPOQPVKNMKLEVKGEFTREVAIPDLRPTAYYIRVRANDPL 412
QY 48 -----SKTG----- 51
Db 413 GPGKLGNOVQIKLPAVPYVNIWEGDEIRVPWMTAFEDCNVTRADVPVLVWLHGR 472
QY 52 ----- 51
Db 473 PLNKGSTQHKMKNGGVLESTQSCVAENEAGSKTKKINVTGTPSAPERIRYQIDGDK 532
QY 60 L-----EILR----- 64
Db 533 VTLQWEPPIITNGPMAGVDVYFTEDPSLPRDQKVHHIDDPNARTTVLRLNEKTPYTFV 592
QY 65 -----QLR----- 67
Db 593 IVGNRLGPGLPSPAPFATTWLAAPVWQLEPSEEMTKEPSNDEMIIECGAGGVKPKI 652
QY 68 ----- 67
Db 653 IWLWSGTLIEDGKEEPRVYDPTTDAQDRTRSKLIAQSTRSGVATCQAVNSEGSEKKV 712

QY 68 -----TTISTKYNVKOPVK----- 82
Db 713 PVKILGPSAPLGIPTPTMHTGFDVAMKPKVTNGRITDYVYVYYSKDPDAPLSDWESKTV 772
QY 83 -----NLTMNTE----- 89
Db 773 PADTRNLTVNVDDEDTPVVVKVQARTDDPGIISAEAYEVTTGRKQVPLSVLEIAPSDVD 832
QY 90 ----- 89
Db 833 PSTGETIVEPTQPIHFRCAVADGRPMPSVYSWLPINASEGDEVPPIPIHSDDSQPHYN 892
QY 90 ----- 89
Db 893 SIQVYSTATTAKRILLCOARNPDGTVDRHVFVNKPGSAPONPEVIVDPNRRVITWQPP 952
QY 90 -----FP-----QYIL----- 96
Db 953 KYPNGEITSNVYITGDESLPVDQWQVFPVDDVTDPKLVLRGALQBPETPFVKIAAVNP 1012
QY 97 -----AGPIQNYSI----- 105
Db 1013 HGEIHTDPKHFDTVSGAPIDAPTDVLPSVSDNTVNITWSPPTQPLGPKSYTYVFAPE 1072
QY 106 ----- 105
Db 1073 YDSDFKTWQRIISVDAPDGDHGEVTLPKGEQFNPTPYKIRISATNDLSEGPASDPVRFE 1132
QY 106 ----- 105
Db 1133 TGSGEIPPTITLDPSTTYVEPLGAATITCTATGVPQPKVHWIKANGETVDSATLQLYD 1192
QY 106 ----- 105
Db 1193 LVKDSATCVAENNACKTQEAWSIQVTPGTAPNEIVLLPMNQENINWETSPDEVNGOI 1252
QY 106 -----TYL----- 108
Db 1253 TNYIIHYGEISDGESEFATWQVTIARDVNHKLANLEPKTYAIRVQAVSDRGPVISA 1312
QY 109 ----- 108
Db 1313 PQVITLPLAPQAITNPIIQVHPNNSVTIEFTPPDDPENPGKKVKDFVIQVTTDEPDDE 1372
QY 109 ----- 108
Db 1373 SVMKELKFTDPDDTDDTTIVSIDGENFNPDTKINTRIIFARGEIDSQNPNEPTLFATGDGVI 1432
QY 109 -----WDFYSTQL----- 117
Db 1433 APSQSFNVDTEDGVRVPAGTDYTIKCVSEGYFADPVVRWVDSHGNSQLSDGFLRLIIR 1492
QY 118 -----RKP-----AK 122
Db 1493 KTLNAKLAENRGGLKETDLTFIVAGPCTAPENIQLTANKETTISVQYEVPSIENGNIISK 1552
QY 123 YV----- 124
Db 1553 YIIYYTLDQDPDQHLGQVQTKPISDWQNVHDMNDGVEGPRKVDIKDFVSTDTAYAVVV 1612
QY 125 -----YS-QY----- 128
Db 1613 QAINDDGPYPYSNQYTIIRTSRAREGPPVELRVPEDCQRSVAQWKEPVTSDVPPIGYEI 1672
QY 129 ----- 128
Db 1673 YVVGDKSVSEDDSAGLNDWIKISIDDDTKLTHKIQLNLLLPDTDYVFKRAIYDPGVSF 1732
QY 129 ----- 128
Db 1733 SEPCIMKTLPGNAPYIQISTGNGVEGSTTIQILPGSQMTACNATGTLPLQVWKIKAG 1792

QY 129 -----NH----- 130
Db 1793 NYEIDPSRVDAGNHAQFSLQVANITEDTTFNCVAQNPLGHANWTINVLIEGLEPNWRD 1852
QY 131 ---TAKT----- 134
Db 1853 DFTVTSKDGQIVLVFNDELPEYLKPPNEWTIQYDDAEQPKDQWESIPSGGAPLTRVEV 1912
QY 135 ----- 134
Db 1913 PNMNGTFFYLVDNPEKGIQTPLVMTPKPPSDIREFGKNDDQIVDFKPAVASEPIK 1972
QY 135 -----ITF 137
Db 1973 EYTSWPESTDPSNVKFTTPADVTSVVVVGLEPDTBYNVQVAABFYEGELASEPVTV 2032
QY 138 RPP-----CGR----- 144
Db 2033 KTFPRDVSCEDHGCAFEMNEDAGTWBPKCYCHGGFHLTSGKSCERDEBDATSOAVLQ 2092
QY 145 ---VPSMT---CLSEMLNVS----- 158
Db 2093 VTPPSITTKVAPEELLTSGSGEVDSTPETLSVVVGPDGKPLVLDKKGKPIDSSGKPKVKFD 2152
QY 159 -----KNDTGE----- 165
Db 2153 ENGDPFAPEGTKLEKNDNGEWYPLVDRNGKPLPVDENNKPITIVIDKGRVVTETDDGT 2212
QY 166 ----- 165
Db 2213 FVTSKGQVEVDDLGRPLDEBGNPKYTNENGQFVLSVDGAVEGDBEEQVQVPLYVD 2272
QY 166 ----- 165
Db 2273 VDDGKYLDEGNEIPVNEGDGPIDVNGKPLEKNEDEKFKVPEKSTQETPQTKITIVSP 2332
QY 166 ----- 165
Db 2333 DGTPLTDASGSPIGLQGPVPTDASGKPLAKDGSPLPTDNNGNVILPSSKNSVDSQPT 2392
QY 166 ----- 165
Db 2393 DDAGRIYVPVLPDGSPLATDSTGNFVNRHGDIVERDDEGKPMGPDGQLLPTDASGNVIY 2452
QY 166 ----- 165
Db 2453 PVTGENDEVLPTDANGNFIYVPVGPDGTPLPTDASGAIVGPDGQPIPTDSNGKPLSKEGY 2512
QY 166 ----- 165
Db 2513 PLPVDNQNYILLPTEIDAAQSLPTDDAGMPYIVKPDGTPLATDSTGSPINDNGEIIIE 2572
QY 166 ----- 165
Db 2573 KDEGRPFGPGLILPTDASGNVIYPAMGLDQQLATDASGNVILVSTEQTWTKSYPVDD 2632
QY 166 ----- 165
Db 2633 SDIITHPIWPDGTPLATDSTGSYVTEGQIIIEKDDEGRPLGPDQVDFQNYEWKLSNC 2692
QY 166 ----- 165
Db 2693 HPLTVLPTDDSGNVIYPVAESGEETKPTDASGKTVYVVRGPDGTPLPTDASGAVIGPDGE 2752
QY 166 ----- 165
Db 2753 VIPTDENGIPLSQDGSPLPTDQNGNVIILVLTSETPTKTLPTIDESGNVVIYPTKPDGTPLA 2812
QY 166 ----- 165
Db 2813 TDSTGSVTEGDTIAKDDDEKPLGPDCEVLPTDASGNVIYVPTVSDQTLPTDDTGKTV 2872
QY 166 ----- 165

Db 2873 YPIRGPDGTPLPTDASGPIVIGPDGEIMPTDENGIPLSKDGTPPLPTDNDGNVIVPSDBET 2932
QY 166 ----- 165
Db 2933 SKELPIDDSGNVIYPTKPDGTPLATDSTGSVTEGDTIIIEKDDEKPLGPDGQLLSTDA 2992
QY 166 ----- 165
Db 2993 SGNVYVDPGLDSQILATTDVYKPIYTVIGPDGTALPTDASGAAGPDGTPISTDETGE 3052
QY 166 ----- 165
Db 3053 LDKGSILPTDDYGNFVFPVVSQELPTDAEVQTPITKPDGTLLATDSSNGVNDNGDIIEK 3112
QY 166 -----OCGNFT----- 172
Db 3113 DDEKPLGPDCEVLPTDGTGNFIYPATTSDEGVIPTDSSGKPLYTIRGPDGTPLPTDTGT 3172
QY 173 ----- 172
Db 3173 SALPGDGEPISTSSGKPLSKDGSPLPTDNNGNVILVPTDESTTKALPTDESNGNVIPI 3232
QY 173 ----- 172
Db 3233 NPDTSLATDSTGSVFVTDGQIIIEKDDEKPLGPDGAILPTDASGNVIYVVPVGPQALP 3292
QY 173 ----- 172
Db 3293 TDETKTVYVVRGPDGTPLPTDASGAVMPDGEPIPTDANGKPLSKDGSPLPTDASGNV 3352
QY 173 ----- 172
Db 3353 LVPSDEVTAKEPLTDESGTIVYVPTRADGTPLATDSTGSVFTDDGQIIIEKDDEKPLGPD 3412
QY 173 ----- 172
Db 3413 GQVLPTDDSGNVIYPAVGPDQAPFTDKSGKPLVVRGPDGTSLPTDASGAAGPDGEVI 3472
QY 173 ----- 172
Db 3473 PTDENGIPLDKGSPLPTDASGNVIYVPSGELTWASHPTDDTGNVIYPTKPDGTLLSTD 3532
QY 173 ----- 172
Db 3533 STGSVFTEGQIIIEKDDEKPLGNGEALPTDGLGNVIYPTDSEQTSPTEDVGTSVHL 3592
QY 173 ----- 172
Db 3593 VRGPDGTPLPTDASGAIGPDGEVIPTDENGVLKDGSPPLTDNNGNVILVPTKESVTK 3652
QY 173 ----- 172
Db 3653 ILPTDDSEAVVHPITRODGTPLSTDTGNTGNFTDNGEIIIEKDDEGRPVGPDQVLTSDVSG 3712
QY 173 ----- 172
Db 3713 NFVYVPTESPNDGKPIHPVLGPDGSPPLTDDSGAVIGPDGEVIPTDASGVPLSKGLPL 3772
QY 173 ----- 172
Db 3773 PTDSGNVILSSDSDVTVELPTDGTGNVIYPTKPDGTPLGTDTGTSFVSDGQIIIEKD 3832
QY 173 ----- 172
Db 3833 DDGKPLGPDGOVLPTDATGNFIYVPLGPDGOALPTDESGKTVYVVRPDPGNPLPTDASGA 3892
QY 173 ----- 172
Db 3893 VIGPGGEPIPTDSSGKPLSADGSPPLTASGNVILVPSDEVTAKESTDESSTIVYVPT 3952
QY 173 ----- 172

Db 3953 ADGTPLATDSTGSGVTDGQTIIEKDEKPLGPDQVLPITYASGNVIYPVIGPDQALPT 4012
QY 173 ----- 172
Db 4013 DESGKTVYVRGPDGTPLPDVSAGAVIGDGEVPTDSNGIPLSQDGTPLPTDNQNGYIL 4072
QY 173 -----TF----- 174
Db 4073 VPTSETATKALPTDESGNVIYPITKADGTPLATDSTGTFVTDGQIIIEKDEKPLGPDG 4132
QY 175 ----- 174
Db 4133 QVLPTDSDSGNVIYPVVGPDGQTDDESKTVYVRGSHPTDGTGNVIYPITKPDGTLATDS 4192
QY 175 ----- 174
Db 4193 TGSFVTEGQIIIEKDEKPLGPDGQVLPDTSRGNVIYVEVKSDEQLLPDTHGTGTVPV 4252
QY 175 ----- 174
Db 4253 HGPDPGLPTDSDGAIIGPDGEVPTDENGIPLSKDSQLPTDNNGNVYLVPSDEGATKT 4312
QY 175 ----- 174
Db 4313 HPTDTSDAVHPITKPDGTPLATDSTGNFVTENGDVITKDEBKPLGPNQILPTDASGN 4372
QY 175 ----- 174
Db 4373 YIYPVIGPDQALPTDESKTVYVRGPDGTPLPDTSAGAVISPDGEVPTDANGIPLDK 4432
QY 175 ----- 174
Db 4433 DGSPLPTDASGNVILVPSBQDITKTLPTDSDGNVIYPITKPDGTSLATDSTGSGFTEGGE 4492
QY 175 ----- 174
Db 4493 IVERDEGKPLGPDGQVLPDTSAGNVIYPVVGPDGQVLPDTDGKTVYVVGPDGIPLST 4552
QY 175 ----- 174
Db 4553 DASGAVIGDGEPIPTDASGRPLDKDGSFLSDASGNVILVPSDAPTNAGVAVVQHQIT 4612
QY 175 -----NPMFFNVP----- 182
Db 4613 RPDGTPLATDSSGHFVTEGDIENDKEGRPIGADGQVLPDTSANNYIFTDVPQTQYAVF 4672
QY 183 ----- 182
Db 4673 IPTDVVPIEAPNCDQDGRVDTLLFVSSHTSAPYLDTLKKLIENLLLTTPRDFLPK 4732
QY 183 ----- 182
Db 4733 IGTLYSATTEITIDISGDPKELFDSTNEIREIGGIPDVTNALRTAKMILETSRGDT 4792
QY 183 -----RWNTKLY----- 189
Db 4793 LVHLHLSFMRTSSKVYTERIRALPNTRLIHLNEKQWAEADPNVALLRSLCIPSEVPLP 4852
QY 190 ----- 189
Db 4853 SNMPTDASGNLLSIPTDEWVTDGTPDDESGFVIYPITKPDGTPLATDSTGSGFVTEGQII 4912
QY 190 -----VGP----- 192
Db 4913 EKNEDGKPLGPDGQVLPDTSNNGNVIYPVVGPDGQALPTDASGKPIYVRGPDGTPLPD 4972
QY 193 ----- 192
Db 4973 SGAVIGDGEPIPTDASGKPLAQDGSPLFVDNEGNYIILPTQQVDTKEYPTDGTGNVIVP 5032
QY 193 -TK----- 194
Db 5033 ITKPDGTLPTDSTGSGFVTENGDRIFNEBKPLGPDGEVLATDASGNVYVPGSVVEPTA 5092

QY 195 -----VNVDSTI----- 202
Db 5093 EPQEVTHGPDGQVLPDTSAGKPIYVRGPDGIPLPDTSAGAAIGPDGDTIATDENGIPLS 5152
QY 203 ----- 202
Db 5153 KDGSPLPTDNTGNVYLVPSDEGATEEKPQSGESIVHPITKPDGTPLATDSTGSGFVTD 5212
QY 203 ----- 202
Db 5213 QVIAKEDGKPIGPDGQVLPDTSNNGNVIYPVIGPDGQALPTDESGKTVYVRGPDGTPLS 5272
QY 203 -----YFL----- 205
Db 5273 TDASGAVIGDGEPIPTDETGLPLNKDGSPLPTDNDGNVILIIPADESVVKALPTDEAKEV 5332
QY 206 ----- 205
Db 5333 YPIVQPDGTPLATDSSGNFVTSSGDIIDIDEGKPLGPDGQALPTDSDGNVIYPVIGPDG 5392
QY 206 -----GLTALLR----- 213
Db 5393 QALPTDESGKTVYPIRGPDGTPLPDTSAGAVIGDGEPIPTDASGKPLSQDGSPLPTDAS 5452
QY 214 ----- 213
Db 5453 GNYILVPSDGEVTKTLPTDDVGNVIYPITKPDGTPLATDSTGSGFVTDGQIIIEKDEGKP 5512
QY 214 ----- 213
Db 5513 LGPDGQVLTSDSDGNVIYPVAVGNGQTIPTDGTGRTVYVRGPDGTPLPDTSAGAVIGPD 5572
QY 214 ----- 213
Db 5573 GEPIPTDASGKPLSADGSPLPTDNNGNVIVPTDGTSTKSHPTDSDGNVIYPVVEDGTP 5632
QY 214 -----YAQR----- 217
Db 5633 LSTDLSGNFLTNSGEIVDRDEKPLGPDGQTLPTDASGNVYVLOKVEETKPLPTDESG 5692
QY 218 ----- 217
Db 5693 NIVVPIPKPDGTPLATDSTGSGFVTEGTVIEKDEKPGVPGDQVLPDTSAGNVIYPDVT 5752
QY 218 ----- 217
Db 5753 PDGQVQPTDVSGKPVYVRGPDGTPLPDTSAGAAIGPDGKPIPTDSNGVPLSEDSGLPT 5812
QY 218 ----- 217
Db 5813 DNQGNVYVPTSETVTYKSMPTDNNRNVIIPTMSDGSLLSTDSGSGFVTEGDKVIEKDE 5872
QY 218 ----- 217
Db 5873 GKPLGPDGQVLPDTSAGNVIYPVHGQDGTPLPTDASGAVIGPDGSLPTDSDGAVIGPDG 5932
QY 218 ----- 217
Db 5933 EVIPTDSNGIPLNKDGLPLSTASGNVIYVSABQGEIEKEIPIITKPDGTLTSDSTGNP 5992
QY 218 ----- 217
Db 5993 ITENGBIIEERDEGKPIGPDGQILPTDASGNVYVYVIGPDGQGLPTDESGKTVYVRGPD 6052
QY 218 ----- 217
Db 6053 GTPLPTDASGAVIGDGEPIPTDASGKPLSQDGSLLPTDNNGNVYVLLPSNEETTQGLT 6112
QY 218 ----- 217
Db 6113 ESNVNIYPITKPDGTPLATDSTGNFVTNGETIEKDEKPIGPDGQTLPTDSDGNVIYP 6172

QY 218 ----- 217
Db 6173 VVPGDQALPTDESGKTIYPVHGPDTPLPTDASGASIGPDGPIPTDSKPLFKOGSP 6232
QY 218 ----- 217
Db 6233 LPTDSNGNFIIVPEKRMDEBLPTDSDSKIIYPIPKPGDTPLASDSTGVFTEDGTIEK 6292
QY 218 ----- 217
Db 6293 DDDGKPLGQVLPPTDASGNVIYPIVPGDGKTQPTDSEKTPYVHGPDTPLPTDASG 6352
QY 218 -----NCHS-- 222
Db 6353 AVIGDGEPIPTDASGKPLSADGSLPTDNNKGYLVPADEVTKVLPTDSDGNVHPIT 6412
QY 223 ----- 222
Db 6413 RPDGTPGTASGSFITDDGOAIEKDDEKPIGPDGQILPIDASGNVIYPIVPGDQALP 6472
QY 223 ----- 222
Db 6473 TDESKTVYVVRGPDGTPPTDASGAVIGLGEPIPTDASGKPLSRNGSLTSSGNYI 6532
QY 223 -----FYLW-- 231
Db 6533 FVPTDDEKSKKCDISSLSLIIIFVLVNDGGAQNYDQFKAVVGFSRKVDNSPDIRL 6592
QY 232 -----NLFRVP-- 237
Db 6593 AVL SVGEIAVPLPLGGYQEKHELSLINSFEIPIVGTETILSPVQAANOQTFPRTGI 6652
QY 238 ----- 237
Db 6653 SKMVVIPADNEEKSTFIIGATYITVKYCTPKDIINTLIEACEKGLVEIVDDTKHVIDE 6712
QY 238 ----- 237
Db 6713 TVPTISSTFVIVDQSGKPLPTDASGNIDNNKPIVIEGEBPTGPDQKLSKNKKGWVY 6772
QY 238 ----- 237
Db 6773 PLVDKFGKPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGN 6832
QY 238 ----- 237
Db 6833 PYKFDNGHVIAPIQIEEBEETTPAIPFIIDGEPINEDDGVYTDKGNVIPTNSEGKI 6892
QY 238 ----- 237
Db 6893 DENGQVLPKNEGBFVKPKADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSNGRP 6952
QY 238 -----KYI-- 240
Db 6953 LAKDGSPLPTDNNGRYVILPSRGYDTEETDESGNVIYPIINPDGTPGTSTGNVITS 7012
QY 241 ----- 240
Db 7013 IGDIIERDDEKPIGPDQVLTPTDASGNVIYVVGPDGLIILPTDATGKIYVVRGPDGTP 7072
QY 241 ----- 240
Db 7073 LPTDASGAVIGNGEPIPTDASGKPLSQDGSPLPTDVNGNYIMLPSDEVTSQSPLTDESG 7132
QY 241 ----- 240
Db 7133 NVIYPIPKDGTPLGTSSGSFITEDQOIEKDEKPIGPDGQILSTDASGNVIYPDVG 7192
QY 241 ----- 240
Db 7193 PDVQTLPTDGDNI SVPTVEATVEFTSDKTPVHHSITKPDGTPSTSTGEFTVEDGQII 7252
QY 241 -----NGTKLK-- 246

Db 7253 EKDEGKPIGPDGQVLPPTDASGNVIYVPIGLDQALPTDKSGKTVYVVRGNGTLPPTDA 7312
QY 247 -----NTM-- 249
Db 7313 SGAIVIGLDGEPIPTDASGKPLSADGSLPTDAVGNVILVPSDDGVIRTHPTDESGNTIYP 7372
QY 250 ----- 249
Db 7373 ITKPDGTPLATDSTGAFVTDGQVIEKDDEKPIGPDGQVLPPTDASGNVIYVPTSSDQV 7432
QY 250 -----RKLKQK-- 256
Db 7433 LPTDABKPVIVDQSGKPLPTDASGNIDNNKPIVIEGEBPTGPDQKLSKNKKGWVY 7492
QY 257 ----- 256
Db 7493 LVDFKGPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGNP 7552
QY 257 -----APVKEQFE-- 264
Db 7553 YKFDNGHVIAPIQIEEBEETTPAIPFIIDGEPINEDDGVYTDKGNVIPTNSEGKI 7612
QY 265 -----KAKKTQST-- 273
Db 7613 ENGQVLPKNEYGEFVKPKADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSSGRPI 7672
QY 274 ----- 273
Db 7673 SKDGSPLPTDASGNVILVPSGEGVTDLSLPTDEAGNIIYPIPKPGDTLLATDSTGSFVADD 7732
QY 274 ----- 273
Db 7733 GQIIIEKDDEKPIGPDGQVLPPTDASGNVIYVPIVPGDQALPTDESGKTVYVVRGPDGTP 7792
QY 274 ----- 273
Db 7793 PTDASGAVIGPDGPIPTDPSGKPLSADGSLPTDINGNVIYVPSDESAAKVLPTDESGS 7852
QY 274 ----- 273
Db 7853 VVYPIPKPGTPLGTASGSFVTDGQAIKGDDEKPIGPDGQTLPIDDSGNVIYVVRGP 7912
QY 274 ----- 273
Db 7913 DQOALPTDESGKTVYVVLGPDGIPPTDASGAVIGPDGPIPTDASGKPLSADGSLPTD 7972
QY 274 ----- 273
Db 7973 NNGNVIYVPADEVTKVLPTDSDGNVHPITRPGDTPGTPLGTASGSFVTDGQAIKODEG 8032
QY 274 ----- 273
Db 8033 KPIGPDGQVLPPTDASGNVIYVPIVPGDQALPTDKSGKTVYVVRGPDGTPSTDASGALIG 8092
QY 274 ----- 273
Db 8093 LDGEPIPTDASGKPLSADGSLPTDAVGNVILVPSDDGVIRTHPTDESGNTIYPIPKPDG 8152
QY 274 ----- 273
Db 8153 TPLATDSTGAFVTDGQVIEKDDEKPIGPDGQVLPPTDASGNVIYVPTSSDQVLPPTDAE 8212
QY 274 ----- 273
Db 8213 KPVIVDQSGKPLPTDASGNVIDNNKPIVIEGEBPTGPDQKLSKNKKGWVYPLVDKFG 8272
QY 274 ----- 273
Db 8273 KPVETDNDKPVITVVDNDGNELSKNDGNWIDLSGNEIDTDELGRPLDSEGNPKFDDN 8332
QY 274 -----TTPYFSY-- 280

Db	8333	GHVIAQIEBEEETTPAIFIIIDGEPINEDDGVYTDKGNVPTNSEKPIDENGQVL	8392
QY	281	-----TTSAAAL-----	286
Db	8393	PKNEDGEFVKPEADTTQSTIVSPDGSPLPTDASGAAIGPDGPIPTDSSGRIPSKDGSP	8452
QY	287	-----	286
Db	8453	LPTDASGNYILVPSGEGVTDLSLPTDEAGNIYPTIKPDGTLATDSTGSAVDGQIIEK	8512
QY	287	-----	286
Db	8513	DDEKPIGPDGQVLPTDASGNIYYPVIGPDGQALPTDESGKTVYVVRGPDGTLPTDASG	8572
QY	287	-----NVT-----	290
Db	8573	AVIGPDGPIPTDPSGKPLSADGSKLPTDINGNYVLVPADEVTKVLPTDSDGNVHPIT	8632
QY	291	-----	290
Db	8633	RPDGTPLGTDSGSFITDQIIVEKNDGKPIGPDGQVLPTDSSDNYIYPSIGSDEQAMP	8692
QY	291	-----NVTYS	295
Db	8693	TDITGSVIVPLVSPDGTVIEGPPKAKVPGDGVKVLPTDASGHFIGPDGPIPTDYGVTVS	8752
QY	296	ITTAARRVST-----STIAY-----	310
Db	8753	DT-----VTTDGC1PLSNDSTGAFITDGTVIENNEDGKPIGPDGQVLPTDAIGNIYPA	8807
QY	311	-----	310
Db	8808	IGPDGQALPTDESGNVYVVRGPDGTLPTDVSAGAVIGPDGPIPTDASGKPLSADGSP	8867
QY	311	-----	310
Db	8868	LPTDNGNVLVPADEVTKVLPTDSDGNVHPITRPDGTPLGTDSGSFVRDDGQAIK	8927
QY	311	-----	310
Db	8928	DDEKPIGPDGQVLPTDASGNIYYPVIGPDGQALPTDESGKTVYVVRGPDGTLPTDASG	8987
QY	311	-----	310
Db	8988	AVIGLGEPIPTDASGKPLSAGSPLPTDNNGNVVLVPADEVTKVLPTDSDGNVHPIT	9047
QY	311	RPD-----SSFWKS-----	319
Db	9048	RPDGTPLGTDSGSFVRDDGQAIKDKGKPIGPDGQVLPTDASGNIYYPVIGPDGQALP	9107
QY	320	-----	319
Db	9108	TDESGKTVYVVRGPDGTLPTDASGAVIGLGEPIPTDASGKPLSAGSPLPTDNNGNVY	9167
QY	320	-----	319
Db	9168	LVAHEVTTKVLPTDSDGNVHPITRPDGTPLGTDSGSFVTDGQAIKDKGKPIGPD	9227
QY	320	-----	319
Db	9228	GQVLPTDASGNIYYPVTSSDQVLPTDAEKPIVDQSGKPLPTDASGNIYIDNNGKPIVE	9287
QY	320	-----	319
Db	9288	GEBPTGPDOKLSKKEKEWYVPLVDFKPGFVETDNDKPIVTVVNDGNELSKNDGNW	9347
QY	320	-----IMATQLRD-----	327
Db	9348	IDLSGNEIDTDLGRPLDSEGNPYKFDNNGHVVIAPQIEEBEATPAIPFIIDGPINE	9407
QY	328	-----	327
Db	9408	DDGVYTDKGNVPTNSEKPIDENGQVLPKNEDGEFVKPEADTTQSTIVSPDGSPLPT	9467
QY	328	-----	327
Db	9468	DASGAAIGPDGPIPTDSNGRPLAKDGSPPLTDNNGRVILPSCRYSGDTETTDESIGNVI	9527
QY	328	-----	327
Db	9528	YPIINPDCTPLGTSTGNYITSIGDIERRDEGKPIGPDGQVLTTDASGNIYYPVVGPDG	9587
QY	328	-----	327
Db	9588	LILPTDATGKIYVVRGPDGTLPTDASGAVIGNGEPIPTDASGKPLSQDGSPLPTDVN	9647
QY	328	-----	327
Db	9648	GNVIMPLSDEVTSQSLPTDESGNVYIPTIKPDGTLGTDSGSPFITEGQIIEKDEGK	9707
QY	328	-----	327
Db	9708	IGPDGQILSTDASGNIYYPDVGPVQTLPTDGMISVPTVEATVEFTSDKTPEVIHSITK	9767
QY	328	-----	327
Db	9768	PDGTLSTDTSGFVTEGQIIEKDEGKPIGPDGQVLPTDASGNIYYPVIGLQALPT	9827
QY	328	-----	327
Db	9828	DKSGKTVYVVRGPNGTPLPTDASGAVIGLGEPIPTDASGKPLSADGSPPLPTDAGNYIL	9887
QY	328	-----LAT-----	330
Db	9888	VPSDDGVIRHTPTDESGNIYIPTIKPDGTLATDSTGAFVTDGQVIEKDEGKPIGPDG	9947
QY	331	-----	330
Db	9948	QVLPTDASGNIYYPVTSSDQVLPTDAEKPIVDQSGKPLPTDASGNIYIDNNGKPIVIEG	10007
QY	331	-----WVY-----	333
Db	10008	EEPTGPDOKLSKKEKEWYVPLVDFKPGFVETDNDKPIVTVVNDGNELSKNDGNWI	10067
QY	334	-----TTLRYRQNPFCBPSPNRRTAVSFMKNTHVLI-----	364
Db	10068	DLSGNEIDTDLGRPLDSEGNFY-----KFDNNGHVVIAPQIEEBEATPAIPF	10116
QY	365	-----	364
Db	10117	IIIDGEPINEDDGVYTDKGNVPTNSEKPIDENGQVLPKNEDGEFVKPEADTTQSTI	10176
QY	365	-----	364
Db	10177	VSPDGSPLTDASGAAIGPDGPIPTDSSGRIPSKDGSPPLTDASGNIYLVPSGEGVTD	10236
QY	365	-----	364
Db	10237	LPTDEAGNIYIPTIKPDGTLATDSTGSAVDGQIIEKDEGKPIGPDGQVLPTDASGN	10296
QY	365	-----	364
Db	10297	YIYPVIGPDGQALPTDESGKTVYVVRGPDGTLPTDASGAVIGPDGPIPTDPSGKPLSA	10356
QY	365	-----	364
Db	10357	DGSPLPTDINGNYVLVPSDESAKVLPTDESGSVVYIPTIKPDGTLGTDSGSGVITEGQ	10416
QY	365	-----	364
Db	10417	LVGKDEGKPVGPQVLPTDSAGHYVYITGADRQILTTDAAGKPIYSVFNEGDIQLPT	10476
QY	365	-----	364
Db	10477	DSSGVAIGHDELVPTESTNGVPLNKDGTPLPTNDSGHFVLVLPATVNDKSKTDEVIVS	10536

QY 365 ----- 364
Db 10537 INPDGTLTGSTGAFVTEDEGPIIENDDGKVPDGOVLPTDSDGNYIYPVIGPDQA 10596
QY 365 ----- 364
Db 10597 LPTDESGKTVPIRGDGTPLPTDASGASIGPDGEPIPTDASGKPLSKDGSPLPTDNDGH 10656
QY 365 ----- 364
Db 10657 YVLVPDDSTIKAFPTDESGNAVYPIPTRDGTPLGTDSGSEFVTDGTIIENDDGKPIG 10716
QY 365 ----- 364
Db 10717 PDGQVLPTDASGNYIYPVIGPDQALPTDESGKTVYPVHGPDGTPLPTDASGAIGPGE 10776
QY 365 ----- 364
Db 10777 PIPTDASGKPLSQDGSALPTDNNGNFILVPSDKSTTKTLPTDESGNPIYPITKPGCVLPA 10836
QY 365 ----- 364
Db 10837 TDSGTNYVTDEGELIEKDNGYPLGDKRVLPTDGSNGVIYPVAGSDEKILPTDNLGKV 10896
QY 365 -----RNE----- 367
Db 10897 YPIPTRDGSPLATDSTGVFVTGDTIVERNEEGKPIGPDQVLTTDSDGNYIYPVIGPDG 10956
QY 368 ----- 367
Db 10957 EPLGTDASGKTVYVRGPDGTPLATDAFNAVIGPDGEPIPTDASGKPLDQSGFPLPTDNN 11016
QY 368 -----TP----- 369
Db 11017 GNYILVPSDEALGKILPTDENGWVSVTNPDGTPLATDSTGSFIASNGLIVERKDEKCP 11076
QY 370 ----- 369
Db 11077 IGPDQVLPTDASGNYIYPVIGPDQALPTDESGKPIYPVFTEGDTQPLPTDSTGFAIGPD 11136
QY 370 -----YTIY----- 373
Db 11137 GELVPTDSANGVPLSKDGSPLPTDASGNYIILPDSGVTTANPTDENGVAIYPIITKPDGTLL 11196
QY 374 ----- 373
Db 11197 ATDSTGSYITQGGQLIEKDNKGKPIGPDQVLPTDGSNGVYYPVVGPDQALPTDDTGNV 11256
QY 374 ----- 373
Db 11257 VYPVINADGSLLATDSSGSFITENGKIVAKDEGKPIPDGQVLPTDASGNYIYPALGPD 11316
QY 374 ----- 373
Db 11317 GSILPTDSNGKSIYPVRGPDGTPLPTDDEGFAIGDGPKEIPTDTSKGPLSADGSPLPTDN 11376
QY 374 ----- 373
Db 11377 NGNYILVLSGVTEHAPTDENGNVIYPVTPNPDGTPLGTDSGAFITQDGTVVKKDEGK 11436
QY 374 ----- 373
Db 11437 IGPDQVLPTDSDGNYIYPVIGPDQVLPTDASGKTVHVSYPDGTQPLPTDASGAIGPD 11496
QY 374 ----- 373
Db 11497 GELVPTDVSGRPLSQDGSPLPTDNNGNVALVVSDEATTKVLPTDEGGNVIYHITKPGSL 11556
QY 374 ----- 373
Db 11557 LGTDASGDFITDHGKAVQKDEGKPIGPDGVLPTDTSNGNYIYPITGPDGNVLPTDSDNGK 11616
QY 374 ----- 373

Db 11617 PVYPVNEDGTQPLPTDSTGSAIDQGBELVSTDSTSGVPLAKDGSPLPTNSAGNYVLVSSG 11676
QY 374 -----GTL----- 376
Db 11677 KSQPTDEHGNVIYPITKPDGTLLATDSTGSYLTEDGQLVEIDDSCKPLGSDQVLPIDAS 11736
QY 377 -----DMSLLY----- 383
Db 11737 GNYIYPALGPDQALPTDDAGNLVYPIVYPDGTPLATESTGNYVTENGVEVVGKNTDGKPI 11796
QY 384 ----- 383
Db 11797 SPDQVLPTDASGNYIYPVAVGPDGOVLPTDASGKLIYPVPHPDGTQPLPTDASGYAVAPDG 11856
QY 384 ----- 383
Db 11857 SLIPTFSGKPLGKDGSLVPTDNSGRYVLVHDDREVTOIPTDESGNTIYPIPTRDGTLL 11916
QY 384 ----- 383
Db 11917 STDSTGIYLTDEGNVIDRDNKGKPLGPDQVLPTDGYGNFVYPADSDIGGAKLLPTDEYG 11976
QY 384 ----- 383
Db 11977 HTLYPVIRPDGSLSTESSGSFVTDGTFVSKSDGKPLGPAGQVLPTDASGNYIYPSIG 12036
QY 384 ----- 383
Db 12037 PDGSPLPTDINGKPAYTVIGRYGDLVPTDSLGRAVINDGSVVPDDEGLPIDOYGVLPT 12096
QY 384 ----- 383
Db 12097 DTTRKCLTLVPTRRPSPFCYVTSIDLILLVIDSSNNIKVLDYRVNMKELIKNPLTEHFNLR 12156
QY 384 ----- 383
Db 12157 KHQVRVGLVKYGDGAEIPVSLGDYDNEDDLVRHISESRLKRAQLGAGLREALDELSIS 12216
QY 384 -----NETMFV-----ENKTASDNKTT 401
Db 12217 GVDGPQVLIVKNGKASDDYSSAVKSLKARNVTVFVVDAGDDDESQQNSLFEEDKTI 12276
QY 402 PTSPSMGFQRTFIDPLWDYL-----DSL 424
Db 12277 VISQWKGADSSVLGPIADYICKIVPNVETSRTPPTKATTSCTGRSCSSIDVESVI 12336
QY 425 LFLDEIRNF----- 433
Db 12337 IVLDSSENFTPDEFVSMKDAVASIVDTGFDLAPDVSKIGFVIYSDKVAVPVALGHYEDKI 12396
QY 434 ----- 433
Db 12397 ELLEKITAEDKINDGVAIALYGLNARQQFOLHGENATKVVILITNGKNGNAAAAAED 12456
QY 434 ----- 433
Db 12457 LRDYGVQLPAVAVGSNPEELATIKLVGNSNTENVIEVAQSTEIDDDAAALLKAVCGNT 12516
QY 434 ----- 433
Db 12517 SPKNSMPAHLTKRDVLAKQFTTAPMLRTTRAVAGGLCNDGIRRPVHFNILVDITSRAS 12576
QY 434 ----- 433
Db 12577 ADEPRRVLHILNFNDRMRDEQHMITINIITVNSKVQNTLSNLRADQLSEQLNAITQQ 12636
QY 434 ----- 433
Db 12637 SDDTVSPKLGAGIDALAELSKENYINGAIKMLIVGSDGTSSDDALPAEYANSDFQHNI 12696
QY 434 ---SLRSP---
|:|

Db 12697 IAVSVRKPATDLKSLKIAGLPTRVVHLDQMSAPNELFDSWIAIYITCDYATASTTRKSTTEK 12756
QY 439 ----- 438
Db 12757 MTTLRPYDRKASKEDATNIELIPLSPSSLSVSWTCTNNKSNYILYTHDTSITKEKWIR 12816
QY 439 -----TYVNLTPPEHRAV----- 452
Db 12817 KEVTCRDSFGTHNELPSDHTYTCVMTNERVDNSTALAIKNCDSLHIDQNTAPEDYV 12876
QY 453 ----- 452
Db 12877 KPSPSSNCOCSEKGAFLRATCEMVIDTNRPIATLPATVDECPCKVKAHGRCRPGYIA 12936
QY 453 -----NLSTNSL----- 460
Db 12937 KDQCQYDIBCEETNNQCSEGCWNTPGSYACACPHGMWRDPLDPFNCVNTANSFKIAAL 12996
QY 461 -----WWL 464
Db 12997 LANYLEANTKNSGSEVTSEKSDGRVNYKATIKSADDKTITPEWSHVPEVVRFAFKWL 13054

RESULT 10
T27935
hypothetical protein ZK617.lb - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27935; T28031
R:White, S.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20442
A:Accession: T27935
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <W12>
A:Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b
A:Experimental source: clone ZK617
R:Harris, B.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z20458
A:Accession: T28031
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7160 <W12>
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b
A:Experimental source: clone ZK629
C:Genetics:
A:Gene: CESP:ZK617.1b
A:Map position: 4
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 593/67/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 19.3%; Score 473; DB 2; Length 7160;
Best Local Similarity 3.1%; Pred No. 2.2;
Matches 217; Conservative 91; Mismatches 130; Indels 6585; Gaps 70;

QY 1 MGRKEMVRDVPKMFVLISIFLLVSFINKVMKSL----- 37
Db 106 VGKPRIIPKGGALIVM-----ECKVKSASTPVAKMKMGVPLSMGGLYHAIFS 154
QY 38 ----- 37
Db 155 DLGQTYLCQLEIRGSSDAGQYRCNIRDQGETNANLALNFEPPDPSRQBRKRSTAS 214
QY 38 ----- 37
Db 215 PRPSSRGSRPSPKSKMSRECTPKRTLKPRGSPSKLRSRTSTPYNVEVSQESRR 274
QY 38 ----- 37
Db 275 SSRTDKMEVDQVSGASKRKPDGLPPPGDEKBLRAGSPSTRKSPRSKASPTPSRKSSA 334

QY 38 -----YNRP-----WR 43
Db 335 GGAASGTTGASASATSATSGSASSDASRDKYTRPPIVLEASRQTRIGGSVVLEVMQ 394
QY 44 -----GLVLSKIGKYK----- 54
Db 395 CHSSTIIEWYRDDGTLVRNSSEYSQSFNGSIKALQVNLKTEESGLYKCHAKCDYEGOSS 454
QY 55 ----- 54
Db 455 AMWKIEQSDVEBELMKHKADEYQKEQKSQTLQAEKTKVARRSKSKSPAPQAKK 514
QY 55 -----LDQL-----KLEILRQL----- 66
Db 515 STTSRGRQEAASEYEHKRSSVRPDPDESQLEIPSSGLTIPERRRELLQGVGESDDE 574
QY 67 -----ETIISTK----- 73
Db 575 VSEISSELPSFAGKPRKTDIFSCVLLYLQSPPKQDMFSDRTLRLKRTTSTKNESSTV 634
QY 74 ----- 73
Db 635 EEKTKLRKTVKVDGELDFKAMVKLKKVKEEGTTEKSGFPLDHDADSTSSVLSQESRSR 694
QY 74 ----- 73
Db 695 RGSNAPPAKDGLEPQANPPAQLKKVKVSGAGGLEKSDSMASLKKLDLKGKIDNDSGAF 754
QY 74 -----YNVSKQPVKNLTWN-----TEF----- 90
Db 755 KVQLKKVVKKEVKESTISVKEKNGTSEGIKTEFMEKREKRTTLQYKEDTSGSKKEDKP 814
QY 91 -----POYY----- 94
Db 815 KKVSIAPVSTNKSDDEPSTPRHHEKVEEKSTSEELKAKVAGVQVQKRNQAOKPEEPKN 874
QY 95 ----- 94
Db 875 LLSQIQLKKVTKKAHDDTNELEGIKLVKVTVPKHVADDDSOSESRRSGSVFGELRGS 934
QY 95 -----ILAGPIONYSI 105
Db 935 RAPDSADNSRRDSIRRSIDMRRESVQELEKSTPLVPAGSAGSAPKIVEP-ENVTV 993
QY 106 -----TYLWF----- 110
Db 994 VENETAILTCKVSGSPAPTFRMFKGSREVISGGRFKHITDGKEHTVALALLKCRSQDEGP 1053
QY 111 -----DF----- 112
Db 1054 YTLTIENVHGTDSADVKLIVTSDNGLDFRAMLKHRESQAGFQKDGEGGGGGGGEKKPM 1113
QY 113 -----YS----- 114
Db 1114 TEARRQSLFPGKVKWKMDIPLPEKTVQOQVDKI CEWKCTYSRPNKIRWYKDRKEIFSG 1173
QY 115 -----TQLRKPAY----- 123
Db 1174 GLKYKIVIEKNVCTLIINNPEVDDTGKTCCEANGVPTHAQLTVLBPPMKYSFLNPLPNTQ 1233
QY 124 ----- 123
Db 1234 EYRTKQAVLTCKVNTPRAPLVYVRGSKAIQEGDPRFIEKDAVGRCTLTIKEVEDDQA 1293
QY 124 -----VYSQY----- 128
Db 1294 EWTARITQDVFSKVQVYVEEPRHTFVVPKMSQKNESDLATILETDVNDKDAEVAWVHDK 1353
QY 129 ----- 128
Db 1354 RIDIDGVKFVKNRKRLLIINGARIEDHGEYKCTTKDRTMAQLIYDAKKNFIVALKD 1413

Qy	129	----	-----NHTAK-----	133
Db	1414	TEVIEKDDVTLMCQHTKTPTGIWFRNGKQISSMPGGKPETQSRNGHTLTKIGKIEMNEA	1473	
Qy	134	----	-----	133
Db	1474	DVVEIDQAGILRGSCNVTVLEAEKKPILNWKPKKIEAKAGBPCVVKVPFQIKGTRRGDPKA	1533	
Qy	134	----	-----	133
Db	1534	QILKNGKPIDBEMRKLVEVIIKDDVAEIVFKNPQLADTGKWALELGNAGTALAPPELFV	1593	
Qy	134	----	-----	133
Db	1594	KDKPKPKPLETKNVNTAEGLDLVWGTPDPDBGAPVKAYIIEMQBSRSGNMAKVGETKGT	1653	
Qy	134	----	-----T	134
Db	1654	DFVKDKLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVPGKPKMMDAIDVDKDHCT	1713	
Qy	135	ITFRPP : :	-----	140
Db	1714	LAMEPPBEDGAPITGYIIERREKSEKDHQVGQTKPCCELTDKKVVEDKEYLYRVKAV	1773	
Qy	141	-----PC----- 	-----	142
Db	1774	NKAGPGDPDCHGKPIKMKAKASPEFTGGGIKDLRLKVGETIKYDVPISGEPLPECLWVV	1833	
Qy	143	-----GRV----- 	-----	145
Db	1834	NGRPLKAVGRVKSSEKHKMIKIENAVRADSGKFTITLKNSSGSCDSTATVTVGRPTP	1893	
Qy	146	----- 	-----PSMTCL- : :	151
Db	1894	PKGPLDIADVCADGATLSWNPDDDDGGDPLTYIVEAQMDNKGKIEVGVKVDNPTTLK	1953	
Qy	152	-----SEMLNVSK----- : :	-----	159
Db	1954	VNGLRNKGYKFRVAVNNEGESEPLSADQYTIKDPWDBPGKGRPEITDFDADRIDIA	2013	
Qy	160	-----	-----	159
Db	2014	WEPHKKGGAPIBEYIIVEVRDPTKEWKVKRVPDNTNASISGLKGEYQFRVRVANKAG	2073	
Qy	160	-----	-----	159
Db	2074	PGQSPSEKOLAKPKFIPAWLKHDNLKSIYVAGATVRHVKIGGEPIVEKWFKNQO	2133	
Qy	160	-----RNDTGE----- : :	-----	165
Db	2134	LENGIQTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADBEKANLTVLDRPSKPNG	2193	
Qy	166	-----	-----	165
Db	2194	PLEVSVDFEDNLNLSWKPPDDGGEPIEYVEVEKLDATATGRWVPCAKVKDTHKAHIGLKK	2253	
Qy	166	-----	-----	165
Db	2254	GQTYQFRVAVNKEGASDALSTDKDTAKNPYDEPGKTGTPTDVVDADRVSLWEWPPKS	2313	
Qy	166	-----QCGNFT----- :	-----	172
Db	2314	DGGAPITQYVIEKKGHGRDQEGCKVSGDQTNAEILGLKEGEYQFRVAVNKAAGPEA	2373	
Qy	173	-----TFNPMFFNVP----- : :	-----	182
Db	2374	SDPSRKVAKPRNLKPWIDREAMKTIITIKGVNDVEFDPVRGEBPPPKKEWIFNEKPVDDQ	2433	
Qy	183	-----	-----	182
Db	2434	KIRIESEDYKTRFVLRGATRKHAGLYTLTATNAGSGDKSHSEVIVLGPSSPLGLEVSN	2493	
Qy	183	-----	-----	182

2494	Db	YVEDRADLEWKVPEDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTPNLPQGHYKPF	2555
183	QY	-----	182
2554	Db	RVRVAVNKEGSDPLTTNTAILAKNPYEVPGVKDPELVDMWDKOHVDLAWNAPDGGAPIE	2613
183	QY	-----RWNTKLYV-----	190
2614	Db	AFVIEKDKNGRWEALWVEGDQKTATVPNLKEGEBYQFRIASARNKAGTDPSPDPSDRVV	2673
191	QY	-----	190
2674	Db	AKPRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPAPDVWTSFNGKGIGESKAQIENBP	2733
191	QY	-----	190
2734	Db	YISRFPALPKARQSGKYTITATNINGDSVTINIKVSKPTKPGPIEVDTVFEDRATL	2793
191	QY	-----GPTKVNVDs-----	199
2794	Db	DMKPPEDDGGEPIDFYEIEIKNTYKDGIIWPCGRSGDTHFTVDSLNGKDHYKFRVKAUNSE	2853
200	QY	-----	199
2854	Db	GPSDPLETETDILAKNPFDRDPRGPEPTDWDSDHVDLKWDPPLSGGAPIEEYQIEKR	2913
200	QY	-----QTIYPLGUTA-----	209
2914	Db	TKYGRWEPAITVPGGQT-----TATVPDLTPNEEYEFRRVAVNKGKSPDPSDASKAVIA	3067
210	QY	-----	209
2968	Db	KPRNLKPHIDRDALKNLTIKAGOSISFDVPVSGEPAPTVTWHWPDNREIRNGGRVKLDNP	3027
210	QY	-----LILLRYAQRNCTHSF-----	223
3028	Db	EYQSLVKVQMERGDSGTPTIKAVNANGEDEATVKINVIDKPTSPNGPLDVSVDHGDHVT	3087
224	QY	-----	223
3088	Db	LNRAPDDDGGIPIENYVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAAV	3147
224	QY	-----	223
3148	Db	NAEGSDPLETFGTTAKDPFDKPGKTNAPEITDMDKDHDVLEWKPPANDGGAPIEYVYV	3207
224	QY	-----	223
3208	Db	EMKDEFPFWDVAHPAGQTNATVGNLKEGSKYEFIRAKNKAGLGDPSDSASAVAKAR	3267
224	QY	-----	223
3268	Db	NVPPFVIDRNSIOBIKYKAGODFSLNPVSGEPTTITWTFEGTPVESDDRMKLNBDGKT	3327
224	QY	-----YLVNAMSRL-----	232
3328	Db	KFHVKRALRSDTGTYIIKAENENGTDTAEVKVTVLDPHPSSPRGPLDVTIVKDGCDLANWK	3387
233	QY	-----	232
3388	Db	EPEDDGGABISHYVIEKQDAATGRWTACGESKDTNPHVDLLTQGHYKFRVKAUNRHGDS	3447
233	QY	-----	232
3448	Db	DPLEAREALIAKPFDRADKPGTPEIVDMDKHADLKWTTPADDGGAPIEGYLVENMRTPS	3507
233	QY	-----LFRV-----PKYI-----	240
3508	Db	GDWVPATVCGELTATVDGLKPGQYQYFRVKALNKAGESTPSDPSRTWAKPRHLAPKI	3567
241	QY	-----NGTKL---K 246	

Db 3568 NRDMFVAQRVAGOTLNFVNVGEPAKIEWFLNGSLSGGNTHIDNNTDNNKLTTK 3627
QY 247 NTMR----- 250
:|
Db 3628 STARADSGKYIVATNSGKDEHEVDVNIIDIPAGEPLRHKDIITKESVYLKWDPLDD 3687
QY 251 -----KLK----- 253
:|
Db 3688 GGSPIITNVVEKQBDGGRWVPCGTSSTLSKVNKLSEGHEYKFRVAVNRQTSAPLTS 3747
QY 254 ----- 253
:|
Db 3748 HAIIVAKNPPDEPDAPTDTVPDWDKOHVDLEWKPPANDGGAPIDAYIVEKKDKFGDWEC 3807
QY 254 -----RKQAP----- 258
:|
Db 3808 ARVDGKTTKATADNLTPGETYQFRVAVNAGKPGKPSDPTGNVAVKPRRMAPKLNLAGLL 3867
QY 259 ----- 258
:|
Db 3868 DLRIKAGTPIKLDIAFEGEPAPVAKWKANDATIDTCARADVNTTPTSSAIHIFSAVRGDT 3927
QY 259 ----- 258
:|
Db 3928 GVYKIIIVENEHGKDTAQCNTVLDVPGTPEGPLKIDIEIHKEGCTLNWKPPDTONGGTDVLH 3987
QY 259 -----VK 260
:|
Db 3988 YIVEKMDTSRGTVQEVGTPDCTAKVNKLVPGEYAFRVAKNVLOGESKPLEAEPIIAK 4047
QY 261 EQF-----EKKAK-----TQSTTTPY 277
:|
Db 4048 NOPDVPDPVKPEVTDWKDRIDIKWNPANNGAPVTGYIVEKKEKSAIWEAGKTP- 4106
QY 278 FSYTTSAAALNVTTNTVY-----SITTAARV--- 303
:|
Db 4107 --GTFPSADNLKGEVEYFRVAVNAGPSDPTDPOITKARYLKPKILTASRKIK 4164
QY 304 -----STSTIAY----- 310
:|
Db 4165 AGFTHNLEVDFIGAPDTATWTGDSGAALAPPELLVDAKSSTISIFPFSKRAADSGNYKL 4224
QY 311 ----- 310
:|
Db 4225 KVNKLGEDEAIFEVIVQDRPSAPGPLEVSDVTKDSCVLNWKPKDGGABEISNYVVEK 4284
QY 311 ----- 310
:|
Db 4285 RDTKNTWVPVSAFVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTEFVLAKDPFG 4344
QY 311 ----- 310
:|
Db 4345 TFGKPRPEIVDTDNHDIDIKWDPDRNGSGPVHDYDIERDKAGTKRWIKVNTSPVQGA 4404
QY 311 ----- 310
:|
Db 4405 FSDTRVQKHTYEYFRVAVNAGKPGQPSDSSAAATAKPMHEAPKFDLDLKGKFRVKAGE 4464
QY 311 -----RPSSTFWKS----- 319
:|
Db 4465 PLVITIPFTASQPDIDISWTKGKPLAGVETDTSQTKLVIPTSTRSDGPVKIKAVNPYG 4524
QY 320 ----- 319
:|
Db 4525 EABANIKITVIDKGPAPENITYPAVSRHTCTLNWDAPKDDGABAGYKIEYQEVGSQIW 4584
QY 320 ----- 319
:|
Db 4585 DKVPLIGSTAYTVRGLEHQOYFRIRAEANAVGLSDYCOQVPVVIKDPDPPGAPSTPE 4644
QY 320 -----IMATQ----- 324
:|
Db 4645 ITGYDTNOVSLAWNPPRDDGGSPLIGYVVERFEKRGGDWAPVQMPVKGTECIVPGLHE 4704

QY 325 ----- 324
:|
Db 4705 NETYQFRVAVNAAGHGEPNSGSEPVTCRPVYVEKPGADAPRVGKITKNSAELTWNRLR 4764
QY 325 -----LRDLA----- 329
:|
Db 4765 DGGAPIDGYIVEKKLGDNDWTRCNDKPEVRDTAFEVKNLGEKEBEYFRVIAVNSAGEPE 4824
QY 330 ----- 329
:|
Db 4825 SKPSDLVLLERQGPPIFDINNLDIIVRAGETIQIRIPYAGGNPKPIIDLFGNSPIFE 4884
QY 330 ----- 329
:|
Db 4885 NERTVVDVNPGBIIVITTTGSKRSDAGPYKISATNKYKGTCKLNVFVLDAPGKPTGPRA 4944
QY 330 -----TW-----VYTTLYR----- 339
:|
Db 4945 TDIQADAMTLSWRPPKONGGDAITNYVVEKRTPGGDWTVVGHVPVGTTLVRNLDANTPYE 5004
QY 340 -----QNPFCEP----- 346
:|
Db 5005 FRVRAENQYGVGEPLETDDDAIVAKNPFDTFGAPGQPEAVETSEAITLOWTRPTSOGAP 5064
QY 347 -----SRNRTA----- 352
:|
Db 5065 IQGYVIEKREVGSTEWTKAAGNILDTHRVGTGLTPKTYEPRVAVNAAGQGEYSVNSV 5124
QY 353 ----- 352
:|
Db 5125 PITADNAPTRPKINMGMLTRDILAYAGERAKILVPPAASPAPKVTFPSKGENKISPTDPRV 5184
QY 353 ----- 352
:|
Db 5185 KVEYSDFLATLTIEKSELTDGGLYFVELENSQSDSASIRLKVVDKPPASQHIRVEDIAP 5244
QY 353 -----VSEFMKQTHVL----- 363
:|
Db 5245 DCCTLVMPSPSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVRNLNTYVGGLIKONRYRF 5304
QY 364 -IRNETPYTI----- 372
:|
Db 5305 RVRAETQYGVSEPCELADVVAKYQFEPVQPEAPTVRDKDSTWAELEWDPDPDGGSKII 5364
QY 373 ----- 372
:|
Db 5365 GYQVQYRDTSSGRWINAKMDLSEQCHARVTGLRQNGEFEFRIIAKNAAGFSKPSPPSERC 5424
QY 373 -----YG----- 374
:|
Db 5425 QLKSRFPFGPPPIHVCAKSIGRHNCTIITWMAPLEDGGSKITGYNVVEIREYGSTLTMTVASD 5484
QY 375 ----- 374
:|
Db 5485 YNVREPEFTVDKLREFNDYFRVAVNAAGKIGPSLPSGPIKIQSSGSGRPOIIVVKPBDT 5544
QY 375 ----- 374
:|
Db 5545 AQPYNRAVFTCEAVGRPEPTARWLRNGRELPESSRYRFEASDGVYKFTIKEVWDIDAGE 5604
QY 375 -TLDMSLLYNETMFV----- 389
:|
Db 5605 YTVVESNPGYSDTATANLVQAPPVIEKDVNPVILPSGDLVRLKLYFSGTAPFRHSLVLN 5664
QY 390 -----ENKTSADSNKTT----- 401
:|
Db 5665 REEIDMDHPTIRIVFDDHILITIPALSVRAGRYEYTVSNDSGEATGTGFWLNVLTGLPEA 5724
QY 402 ----- 401
:|
Db 5725 POGPLHISNIGPSTATLSWRPPVTDGGSKITSYVVEKRDLSKDEWTVTVSNKMNVIYT 5784

QY 402 -----PTSP-----SMGFO 410
DB 5785 GUFENHEVEFRVAQENIGAPLVSEHPHIIARLPDPPTSPLEINLEIVQGVGYVTLNQ 5844
QY 411 RTFIDPLWD-----
DB 5845 R-----PLSDGGRLRGYIVKOEHEDEWFRCNQPNPNVNVNPLIDGRKYRVFVAV 5900
QY 420 -----YLDLLF-----
DB 5901 NDAGLSDLAELDQTLFOASGEGPKIVSPLSLENEEVCVTFECBISGSPRPRWFK 5960
QY 427 -----
DB 5961 GCKELVDTSKYTLINKGDKVLIINDLTSDDADEYTCRATNSSGTRSTRANLRKTKPRV 6020
QY 427 -----
DB 6021 FIPPKYHGYEAQGETIELKIPYKAYPQGEARWTKDGEIENNSKFSITTDKDFATLRI 6080
QY 427 -----LDEI-----
DB 6081 SNASREDYCYRVVVVENSVDGTVNVTADVPEPRPPIIENILDEAVILSWKPPALD 6140
QY 431 -----
DB 6141 GGLSVNTYTIKREAMGGSWSPCAKSRYYTTTIEGLRAGQYEFRIIAENKHGQSKPCP 6200
QY 431 -----
DB 6201 TAPVLPDGRKRRGYVDDEGKIVRGKTVSSNYDYNVDFWQYVQPVVEIKHDVL 6260
QY 431 -----RNFS-----LRSPY 440
DB 6261 DHDYIHELGTGAPGVVHRVTERATGNFNAKVFVTPHESDKETVKIKIQTMSVLRHTL 6320
QY 441 VNL-----
DB 6321 VNLHDAFEDDNEMVMIYBFMSGGELPEKVADEHNKMSDEAVEYMRQVCKGLCHMHENY 6380
QY 444 -----
DB 6381 VHLDLKPENIMPTTKRSNELKLIIDGLTAHLDPKQSVKVTGTABFAAPEVAEGKPVGY 6440
QY 444 -----
DB 6441 TDMNSGVLYLLSGLSPFGGNDDETLRNKSCDNWMDSAFSGISEDGKDFIRKLL 6500
QY 444 -----TTP-----
DB 6501 ADPNTRMTIHOALEHPMLTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGRIS 6560
QY 447 -----EHR-----
DB 6561 NYSSLRKHPOEYSIRDAFWRSEAOPRFVVKPYCTEVEGOSANFYCRVIASSPPVTV 6620
QY 450 -----RAVN-----
DB 6621 HKDBELKQSVKMYKRYNDYGLTINRVKGDGKEVTVRAKNSYGTKEBIVFLNVRHS 6680
QY 454 -----
DB 6681 EPLKEPLEPMKAPSPRVEBEFKERRSAPPTFHLNRNLIQKHOCKLTCSLQGNPNT 6740
QY 454 -----LST-----
DB 6741 IEMKDGHPVDEDRVQVSRGVCSLIEFNARVODAGTYTATNDLGVDSVSECVLTQ 6800
QY 454 -----
DB 6801 KGGEPIRVSSFRPRRAYDTLSTGTDVSHSYADMRRLSIRDVSPDVRSAADDLTKI 6860
QY 457 -----

DB 6861 TNELPSFTAQLSDSETVEGSAEFAAVSGQPEPLIEMHNGERISESDSRFRASYVACK 6920
QY 457 -----SNS-----
DB 6921 ATLRISSDAKSDGQYLCRASNSAQEQTRATLTLYKGDOPLNGHAGQAVESELRVTKHL 6980
QY 460 -----LM-----
DB 6981 GGEIVNNGESVTFEARVQGTPEEVLMMRNGOBLTNGDKTISQDGETLSFTINSADASA 7040
QY 462 -----
DB 7041 GHYQLEVRSGKTNLVSVASLVVVGKADPPVTRLPSSVAPLGSGTAFTIEPENVEGLTV 7100
QY 462 -WW 463
DB 7101 QWF 7103

RESULT 11

T32650

hypotheical protein F39C12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C:Accession: T32650

R:Chisoso, S.; Sansone, J.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F39C12.

A:Reference number: Z21206

A:Accession: T32650

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-5105 <CHI>

A:Cross-references: UNIPROT:O61201; EMBL:AF039043; PIDN:AAB94194.1; GSPDB:GN00028; CESP

A:Experimental source: strain Bristol N2; clone F39C12

C:Genetics:

A:Gene: CESP:F39C12.1

A:Map position: X

A:Introns: 19/2; 47/3; 103/2; 142/2; 174/2; 324/2; 362/3; 436/3; 494/1; 523/1; 923/3; 98

88/3; 4971/3; 5020/3

Query Match

Best Local Similarity 19.3%; Score 472; DB 2; Length 5105;

Matches 213; Conservative 94; Mismatches 141; Indels 4284; Gaps 66;

QY 1 MGRKEMM-----VRDVPKM-- 14

DB 319 MLRRKMPYEDAEVPLLGYYGYEAAKNDPOVPIPTLWTPAEVDMKFLAKIQDVPEIDY 378

QY 15 -----FVLISISFLVSPINC----- 30

DB 379 ESYLYKAVERSAERDPKIHKPKILYQAKGVESPDFV-----PFCVNGIMCFDRNTFTYK 433

QY 31 ----- 30

DB 434 IVRSKPKHYGSPNPYAGSDSVFFYRGLQIQEMIPAQTLLTMNCRLEAGNAYQVRSERSLK 493

QY 31 -----KVMKSKALYN----- 39

DB 494 ELAPFKITEERSVWAKWLHNSVIIQYQGEKFGNIGVNDLEWEDLVLPDGLPEVEIPPY 553

QY 40 -----RP----- 41

DB 554 EYMKONVATEITOMVAPFFFLPDPRLDLSSESETHNAPPNGNCEDDSDEPESG 613

QY 42 -----WR-- 43

DB 614 LPSELSFNMRTCKCHAKIRYNNKQSDRKVMYASMGNSFAIGSKRIRDKFFKLGWKS 673

QY 44 ----- 43

DB 674 TKPQILGHLRKIDTAFKKKSIDLDPKQRFKPRBFQKRVLPKLTIDTSVFMKPEYIK 733

QY 44 -----GLVLSK----- 49
 Db 734 EREAKEKEMREKAAOPTQAIQIGLSMPKSRKEPKQTPVTSNGNEEDPEHARPMIRAAIQ 793
 QY 50 ----- 49
 Db 794 LADNNSDDDEIVSPQATPTPPAPTPTPHVPOALSVRNEKDDPEYETPEDAYFGIKKR 853
 QY 50 -----IGVKLDOLK----- 59
 Db 854 MYVYQATOPQLGLRRYKPDRLRESTFIRLEHLAMYGRRRHERDILYGRFEENDKQSS 913
 QY 60 -----LEILROL-----ETT 69
 Db 914 VKERCTLDLKFSLNNHMKVQOTFYDKNPFPSNKCYSNQLIRKAIQMYNRLCGFDSHET 973
 QY 70 I----- 70
 Db 974 FKSVRHERMLCGVPRNVMDKEAIEEARYLKDISIENSTIKLYNHFTFVQAOLPYKL 1033
 QY 71 -----STKN----- 75
 Db 1034 DHTIFMENEGDTPFSQABCEPIEEKQTAKYDENDKFMPLGLRLRPFCEFDIPPFMHEI 1093
 QY 76 ---VSKQPVKN----- 83
 Db 1094 ENRRVPKPKNPGLYLEFHLSAQPSKPVLRSSNSCPDLGLAIESTENSVCNEFFHE 1153
 QY 84 -----LTWNT 88
 Db 1154 IYRRIIAYDMCHQREHGNVPHYGSEYLEPLLTAGVLOQFNTHSVNGDYKVDDILTNTPT 1213
 QY 89 EF-----PO-----YYILAGPIQNY 104
 Db 1214 KLEWDSPMISDTRDTGRKFMRRVVPFKEPOPPROLEGIDPLFDYITIEVNRAAKNLH 1273
 QY 105 IYLMFD-----F 112
 Db 1274 VSVIFNDPILRTWPGILDEGRGMFTLCPPRPVPABQLSEIDI PKSPCLPTKAVEF 1333
 QY 113 YS-----TQLRKPQKVYQYNHTA----- 132
 Db 1334 YSDFTEVLLIGRDSKUNIRPVPVNEHNINVEGLIEILKPAFYGFYKPYDRESYNEEQ 1393
 QY 133 -----KTIIFRPPPCGRVPSMTCLSEMLNVSK----- 159
 Db 1394 SLEKNRRELQRAIFEKRIITFRQVIBETEQLAIEENLNVEEDDLRTTSESEGEDEV 1453
 QY 160 RNDT-----GRQG----- 167
 Db 1454 KNDSDPNDISDFWKIFKVTEDEAGVNGDRDSDDELVORKIGIFHAGDDSYNCFNPHTRE 1513
 QY 168 ----- 167
 Db 1514 NIRKISVBEVEDEBPKLGSVLTNLFEQDLVKQRKLGDCPIGSPRKSSI KIDYNRGMV 1573
 QY 168 ----- 167
 Db 1574 PLTELTKQTKVTWGEVTEHSVPPRPPIRHRNRHRVEVDKFPVLDYLSVOLTRRVKNSDNF 1633
 QY 168 ----- 167
 Db 1634 QXTKPKRNSDPTAVHHIGIPRRASLDQFYRKDLPATCSACIKEPFWPTKEICIVCHPDTT 1693
 QY 168 -----CGN----- 170
 Db 1694 IFRCTNIVNLHEPCLICGNNIYKFSNVEHNNQVVTPIISIFDQPKLPVDEEFKKLA 1753
 QY 171 ----- 170
 Db 1754 VKNETVIYENTPAVMNVKAYRAGESAVNTFFAETDDRGFKTLQGDLAGFPHREFNMTW 1813

QY 171 -----FTTF----- 174
 Db 1814 RMEDEKSERVFVSEFDFVAFDRDADIDMWRDTMTNISQDGEWDMWTTDYDQRTGMRIRRWI 1873
 QY 175 ----- 174
 Db 1874 SYAALATRVQKQNEKVHYRKKAAEREARILSGEISVPSLNLNELKKAKOKEVAATIATE 1933
 QY 175 ----- 174
 Db 1934 NDDAKKEERRKAYKWKLPPIPEFIEDPKKKNFAPPKSEDMLESSGSEDELPPIKSR 1993
 QY 175 ----- 174
 Db 1994 ILSKEMKEFIDTIRVHMEQKPSFEGIAVTYGLFEKRPILKNDIVLVRKKNVRKPPPP 2053
 QY 175 -----NMFNFVPRWNT----- 186
 Db 2054 FRIQRIPLDESMLNATFTVSPFVPVQVRNPRDPMKWTVPFYNQSORLALTAQETKERALR 2113
 QY 187 -----KUYV-----GPT 193
 Db 2114 PPTPPPTLKKLYTTMKDVEKAFYHRKRVKSOITLLKAKTARQLRHFHKLYGHSKEPSHPT 2173
 QY 194 KV----- 195
 Db 2174 KAWTHVHRQIGLKKSVQPRKLEAHYLPGLAKOKNKYIPTYFPNPKLEKVKLSQSKPSR 2233
 QY 196 ----- 195
 Db 2234 ISSASDSEYCPIVKFPVKQKRVVELKEDHPNTNSELGI IERKKELKREBPPAVLMSSTR 2293
 QY 196 -----NVDSQT----- 201
 Db 2294 STLKRPETGFSVGREISVGRKPALPKADVQKLTETVSLKRIDSQTEPKTESIKORRHIFE 2353
 QY 202 -----IYFLGLTAL----- 210
 Db 2354 HLEDKSNINFSQIGIPALISVVEPSSCOLPKRSQKEHKITTNSSNSISIDSGMKMDRSS 2413
 QY 211 -----LLRYAQNC----- 219
 Db 2414 KQPSVAVRSTWIMPHSRKLRKQMPACRSPSVDKGSGYRKALKIARKCRKSIQKEKSKV 2473
 QY 220 ----- 219
 Db 2474 TKPDDSKPEVIPSQKPKSQLETSVALPSKSAKVSTKSEKOLSPKSTELQERKFLKN 2533
 QY 220 -----THSFYL----- 225
 Db 2534 KVHNRSSRAPGIHHQRPPLHPRMLPTDHTDTRKDSRAASQFRFQHMRSKRFTQKVLHK 2593
 QY 226 -----VNAMSRNLFRVP-----KYIN 241
 Db 2594 EPRRSVLNITIRKNDLSLKGKHNRMKYEQPIDAKISDMFLKPMKPLRRSKGIRSKKYK 2653
 QY 242 GTKLKNT-----MRK----- 251
 Db 2654 NVKAKNTTKLOPMWKQMLRKLADYNNVIRKRVLSDFKIRPKQOEDMETKEKSIIVVK 2713
 QY 252 -----LKRQO-----APVKEQF-- 263
 Db 2714 PSNNLKSCLKRGQGHRSKMFNSQSPDGKTSAGSNKTLKLVKTSAGHLAGTSPPKTSFTV 2773
 QY 264 -----EKKAKK----- 269
 Db 2774 PSILLEDDSSPLKNVTLNTRKNSKGEKVKENTAPVQOAPVPVSPPRETPEPTSPVP 2833
 QY 270 -----TOSTTTP----- 276
 Db 2834 RLLRPLSABSEKVPVEKVDTPSLTPFPAASVPAPPPMPKPADQOQWRAAMWARRYMCAKSP 2893
 QY 277 -----YFSYTT----- 282

Db 2894 VFVKPSALSATESLKEASNTQESASTTRVQPALVSHSTPEPVQLTRMAAKKSEQKEY 2953
QY 283 -----SAAALNV----- 288
Db 2954 KTPVAKPKATAPPITPKAASIRPATQKLHESELTKSEASRNVSSSSGSKSQKQADVQSRP 3013
QY 289 -----TTNVTV----- 294
Db 3014 KLSQVQEKRSVPKVVKPLDPKPGQKEQESKNEKVVPSNTVTLDDQASSNSNQQPGSAK 3073
QY 295 -----ARRVST----- 307
Db 3074 KPSASIFTALPGRKKEKMTSSTAESSIPIPSPENQEAETSNRSPILPRIKPKETGQ 3133
QY 308 ----- 307
Db 3134 LIKESPTSDVPKPCVVSQAIVKFDGKPISNSPSYKGMARRRRPKADSAADSGDLGDE 3193
QY 308 -----IAYRP----- 312
Db 3194 GSTSHEGCLKDENSQKPKAGKTVRFGPISEYEPPEPAPPQVEIPKVEPTPEISFESLAQ 3253
QY 313 ----- 312
Db 3254 QMAQKTLRKEASKILEHPVTEQTSPTABVPSFTKSGSKLSPFGLSPFGKVTPKQNTA 3313
QY 313 ----- 312
Db 3314 QTQSLTVLESTSSPELLTPSNPENTASTSSSKQPKSLSEBKITATQMIQTRGRPIRKL 3373
QY 313 -----DSS----- 315
Db 3374 SHRQLKIGDPSPLQLSGDSSLPRAPQSPKPPQLEVPGVQVQYNNPPSQRYVTLRGRE 3433
QY 316 ----- 315
Db 3434 WKPPPKAKTPLSIRVQSGVSTSPAILDSAKEVLLAPIEKTQKQKRPKPMETTLVLVD 3493
QY 316 ----- 315
Db 3494 GTEHFKKLTAEQLLSRQFKSKRKNKKTWTWSVIGCDENDDSTKQMHKLLBFRD 3553
QY 316 FMKSMATOL----- 325
Db 3554 LKKSLLITRQDKMITARAVFNPRFTDSSSTKHELLRNKPKFETTRKKLKLANNRQKI 3613
QY 326 -----RDL-----ATWVY- 333
Db 3614 RMSGIHQPKRDVAISAIROVQKVRFFQSTGKKMMNPIISGFKPKSILKPPKPPATWLF 3673
QY 334 ----- 333
Db 3674 IDSEGNKLYMKLGPQDWQKNDKCNPMGANDPRKKKSNSYKRWAKOMAKVQKQNTK 3733
QY 334 ----- 333
Db 3734 LFETNQLTRRLGAPKDSHPKDLVDPKADAPPPSRSLSWTTVSSNASLSLLDALNTV 3793
QY 334 ----- 333
Db 3794 GSDEPSTSGIVTDIAPPIVKEQFRKAKESRPTQQTVSQKYPTVYLGMTVARSPSSSTSL 3853
QY 334 -----TTLRYONPCEP----- 346
Db 3854 VNLDERAVSSLAODGSEIVEYVQEBPVVETTKRRFRDFVRPISALFKAFPTSSSPEVGL 3913
QY 347 -----SRNR-----TAVS----- 354
Db 3914 EADSRSPRTERTLGLFRSSPTMPEFSTRHVSSSPTERRANRGPVTSVQGSSEADILA 3973
QY 355 ----- 354

RESULT 12
I38346

Db 3974 LMOAQPDALENMBEQVVSKEPPIQLQAPVLNKGSMWRPKLKLKNSLYTQFKELKEESKUR 4033
QY 355 -----EFMK-----NTHVLIRNETPYTYITGLD 377
Db 4034 KLVEYLDQNYKNNKSYSTLQTLITFNLDQYMRTNLYPRPKLNTDLL-----GKYD 4083
QY 378 MSSL----- 381
Db 4084 PNSLRVSSSSRGLQRSTRLIVKKIRRCQBLEKWKELIATSTNQQTNGELVKRAKAS 4143
QY 382 ----- 381
Db 4144 QKASNRKAKKVLVLYKFKRTCKFRVQVTTKSGARKILADONQKKQLIMHEEKRTPTIL 4203
QY 382 ----- 381
Db 4204 TIGDEDDENEIVDLLSGNEDSANQNLADYDGDNDDDAIDTELDWVANTSQVIRKQRR 4263
QY 382 -----YVN----- 384
Db 4264 RLLRHYFNFAIRMARKVSRRSVGQKPLVFYSFISQMSRMRNPMRIRHNDQTRARQNK 4323
QY 385 -----ETMFVENKTASDSNKTTPSPS-----NGFQ 410
Db 4324 LAVSNSTTLKETLRSSSKQNRKSKSQROIPIICIRNSLYGCTABKQVIVEQSKFNQ 4383
QY 411 R-----TF-----IDPLM----- 418
Db 4384 KLATKLTQTFECLDEYIDVCTPIKLVPSKNTILPKQWFDVKTCLIDPVYVTSPCNTD 4443
QY 419 -----DYLDSLL-----FLDE----- 429
Db 4444 SDFSDNTDDYKSDKQLFRESLESTYHVKGGLGLYLARLVQERVADDVERNIIPRG 4503
QY 430 -----IRNFSLRSP----- 438
Db 4504 PGMETCNLVVYMTPIRRLMLNSFGLRMPDRLLDVMLSKHQNEKPELFDRLYMLCYGRG 4563
QY 439 -----TY----- 440
Db 4564 DLCKYIIDKISLQGPNNKYEKMDITYQAHIIHKLYKQMYIENGREAKLAVQEFMDIG 4623
QY 441 ----- 440
Db 4624 LDIDHEPFSQVQWQYDRMLYGNPKPNKDNWVEACDQFLATYNTKEDVLVSCQASILL 4683
QY 441 -----VNLTPPEH----- 448
Db 4684 MRMIPLGGYSQDVVDHLITWNLPMRDKLFIINLTPPDNLVENIVKIFHTVVRGLDRLPFV 4743
QY 449 -----RRA----- 451
Db 4744 GLGYHADPLRVLFEEAVESGDPQLIAHIVRGRCLDRSRPEMWTDCGDHIQKDFANI 4803
QY 452 ----- 451
Db 4804 MTKVITNGNERRPYGAPTMDEMPFRYVRNPDGSDVAETTEELANIMECEVTRYMAILEL 4863
QY 452 -----VNLSTNSL----- 460
Db 4864 TKGAAVTKGDLDNHPKNTYPHPKPYFDVLCTGCKMNVSKGRNLGRGLRTLCVSTMPLEIQ 4923
QY 461 -----W----- 461
Db 4924 DKDKKRPNAKTMSPRLTPENYIETLDPTEKMLMAKIDBEEKTDITVQVLPSPFGCFQC 4983
QY 462 ----- 463
Db 4984 RPSLPRCCVCMCPYLNHDHMDFALKDYNEPSQFSICNICNHGGHVNHIATWF 5035
-----W 463
----- 463

elastic titin - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
 C:Accession: I38346
 R:Label: S.; Kolmerer, B.
 Science 270, 293-296, 1995
 A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
 A:Reference number: A57430; MUID:96026330; PMID:7569978
 A:Accession: I38346
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-7962 <RES>
 A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g1017426
 C:Genetics:
 A:Gene: GDB:TTN
 A:Cross-references: GDB:127867; OMIM:188840
 A:Map position: 2q31-2q31

Query Match 19.2%; Score 471; DB 2; Length 7962;
 Best Local Similarity 2.7%; Pred. No. 3.7;
 Matches 211; Conservative 91; Mismatches 154; Indels 7235; Gaps 68;

QY	2 GR-----	3
Db	158 GREIYSDKCSIRSKYISSLEILRTQVDCGYCTKASNEYGSVCTATLTVAEYPT 217	
QY	4 -----	3
Db	218 FLRPKSLTTFVGKAAFICTVGTPTVETIQKDGAAALSPSPNWFISAENKHILELSN 277	
QY	4 -----KEM-----	6
Db	278 LTTQDRGVSKASNKFGADICQAEIILDKPHFIKELEPQVQSAINKVHLCQVDEDRK 337	
QY	7 -----M 7	
Db	338 VTTWSKDGKLPKDYKICPEDKIATLEIPLAKLKDSGTVCTASNEAGSSSCATVT 397	
QY	8 VRDVPKMFVLSIFLLV-----	25
Db	398 VREPPSFVKVDPDSYLMPLPESARLHCKLKGSPVQVTFWFKNNKELSESNTVMYFVNSE 457	
QY	26 -----SPT-----	28
Db	458 AILDITDKVEDSGSYSCAVNDVSGSDSCSTEIVIKPEPPFIKLEPADIVRGTNALLQC 517	
QY	29 -----NC 30	
Db	518 EYSGTGPEISWFKDKQIRSSKRYLFSQSLVCLIEFNSADVGVECVVANEVCKC 577	
QY	31 KWSKALYNRP-----	41
Db	578 GCMATHLLKEPPTFVKVDDLIALLGGQVTLQAAVRSSEPISTVMKRGQVIREDEGKIM 637	
QY	42 -----	41
Db	638 SPNSGVAVLIIPDVQISFGKYTCIAENEAGSQTSVGLIYKEPAKIIERAEILQVTAGD 697	
QY	42 -----W-----RGVLISKIKYKLD-----OLK-----	59
Db	698 PATLEYTVAGTFELKPKRYKDGRLVASK--KYRISFKNVNAQLKPYSAELHDSGQVTFE 755	
QY	60 -----	59
Db	756 ISNEVGSSSCETTFTVLDRDIAFFTFLRNVDVSVNGTCRLDCKIAGSLPMRVSWFKDG 815	
QY	60 -----LEILR-----	64
Db	816 KEIAASDRYIAFVEGTASLEIIRVDMNDAGNFTCRATNSVGSKDSSGALLVQEPSPFT 875	
QY	65 -----	64
Db	876 KPGSKDVLPGSAVCLKSTFGQSTPLTIRWFKGNKELVSGGSCYITKEALESLELYLVT 935	

QY	65 -----QL-----	66
Db	936 SDSGTYTCKVSNVAGVECSANLFVKEPATFVEKLEPSQLLKGDATQLACKVTGTPTPIK 995	
QY	67 -----	66
Db	996 ITWPFANDREIKESSKHRMSFVESTAVLRLTDVGIEDSEYMCENAGSDHCSSIVIVK 1055	
QY	67 -----	66
Db	1056 ESPFTKEFKPIEVLKEYDVNMLAEVAGTPPEITWFKDNTILASGRKYKTFIQDHLVSL 1115	
QY	67 -----	66
Db	1116 QILKFVAADAGEYOCRVINEVGSSICSAVTLRPPSPFIKIESTSSLRGGTAFAQATLK 1175	
QY	67 -----	66
Db	1176 GSLPITVTLKDSDEITDDNIRMTFENNVAISLYLGLIEVKGDKYVCQAKNDAGIQRC 1235	
QY	67 -----ETISTKYNVS--	77
Db	1236 ALLSVKEPATITEANVSIDVTQGDPAQLVKFGTKEITAKWPKDQOELTLGSKYKISVT 1295	
QY	78 -----KOPVKNLTM-----	86
Db	1296 DTVSILKIIISTEKKSGEYTFEVDVGRSSCKARINVLDLIIPSPFTKKLKWDSIKGS 1355	
QY	87 -----NTEP-----	90
Db	1356 FIDLECIVAGSHPISTQWFKDDQEIASEKYKPSFHDNTAFLEISOLEGTDSTGYTCSAT 1415	
QY	91 -----POYYI-----	95
Db	1416 NKAGHQCQSHLTVEKPPYFVEKPSQSDVNPTRVOLKALVGTAFTMTIKWPKDNKELHS 1475	
QY	96 -----	95
Db	1476 GAARSVWKDDTSTSLBELFAAKATDSGTTCQLSNDVGTATSKATLTVKPPQFIKKPSPV 1535	
QY	96 -----	95
Db	1536 LVLRNQSTTFEQINGTPKIRVSWLDGNEITAIQKHGISFIDGLATFQISGARVENS 1595	
QY	96 -----	95
Db	1596 TVVCEARNAGTASCSIELKVKKEPPTFIRELKPVEVVKYSDVELECEVTGTPTPEVTWLK 1655	
QY	96 -----	95
Db	1656 NNREIRSSKKYTLTDRVSFVNLHITKCDPSDTGTCVQSVNEGGSCSSTRVALKBPSPF 1715	
QY	96 -----LAG--PIQNYGITYL-----	108
Db	1716 IKKIENNTTLKSSATFQSTVAGSPPI--SITWLKDDQILDEDDNVVISFVDSVATLQI 1772	
QY	109 -----	108
Db	1773 RSDNGHSGRYTCQAKNESGVERCVAFLLVOEPAQIVEAKSVVDVTEKDPMTLECVVAGT 1832	
QY	109 -----	108
Db	1833 PELKVKWLKDGQIVPSRYFSMSPENNVAISFRIOSVMKQDSGOYTFKVENDFGSSSCDAY 1892	
QY	109 -----	108
Db	1893 LRVLQDNIPTFTKLTMDKVLGSSIHMECKVSGSLPISAQWFKDGKEISTSAKYRLVC 1952	
QY	109 -----	108
Db	1953 HERSVLSLVNLELEDTANVTCKVSNVAGDDACSGILTVKEPSPFLVKPQQAIPDSTV 2012	

QY 109 -----WF-----DFYST----- 115
Db 2013 EFKAILKGTPEFKIKWFKDDVELVSGPKFCIGLEGSTSFNLNLYSDVASKTGOYCHVTND 2072
QY 116 ----- 115
Db 2073 VGSDDCTTLLVTEPPKFKLEASKIVKAGDSSRLECKIAGSPEIRVVWFRNEHELPA 2132
QY 116 ----- 115
Db 2133 DKYRMTFIDSVAVIQMNNLSTEDSGDFICEAQNAGSTSCSTKVIVKEPPVFSSFPPIVE 2192
QY 116 -----OLAKPAKYYSQYN-HTA----- 132
Db 2193 TLKNAEVSLECELSGTPPEVVVWYKDKRQLRSKKYKIASKNPHTSIHILNVD7SDIGEY 2252
QY 133 ----- 132
Db 2253 HCKAQNEVGSDTCVCTVKLEPPRPFVSKLNSLTVVAGEPABEQASIEGAQPIFVQWLKEK 2312
QY 133 ----- 132
Db 2313 EEVIRESENIRITFVENVATLQAKAEPANAGKYICQIKNDGGMRENMATLWLEPAVIV 2372
QY 133 ----- 132
Db 2373 EKAGPMTVTVGCTCTLECKVAGTPBELSVEMYKDGKLLTSSQHKRFSFYNNKISSLRILSVE 2432
QY 133 ----- 132
Db 2433 QODAGTYTFQONNVGKSCTAVDVSDRAVPPFTRRLKNTGGVLGASCILECKVAGSS 2492
QY 133 ----- 132
Db 2493 PISVAMFHEKTIKVSAGKYQTTFSNDVCTLQLNSLSDSDMGNYTCVAANVAGSDECRV 2552
QY 133 -----KTIIFR-----PPP----- 141
Db 2553 TVOBPPSFVKBPBPELVLPKGNVTFTSVIRGTPPFKVMWFRGARELVKGDRCNIYFEDTV 2612
QY 142 ----- 141
Db 2613 ABLELFINIDISQSGEYTCVSNNAQASCTRLFKPEPAFLKRLSDHSVEPGKSIILES 2672
QY 142 -----CGRVP-----SMTCLSEMLNYSKR-----NDTGEQ 166
Db 2673 TYTGLPISVTWKDGFNITTSKCNIVTTEKTCILEILNSTKRDAGQYSCEIENAGRD 2732
QY 167 GCGNF-TTFNPMFP-----NVPWNTKLYVGPT----- 193
Db 2733 VCGALVSTLEPPYFVTELEPLEAAVGDVSLQCQVAGTPEITVSWYKGDTKLRPTPEYRT 2792
QY 194 -----KUNVD-----SOTIY----- 203
Db 2793 YFTNNVATLVFNKNVINDSGBYTCKAENSIGTASSKTVFRIQERQLPPSFARQLKDIEQT 2852
QY 204 ----- 203
Db 2853 VGLPVTLCRLNGSAPIQVCHYRDGVLRLDHENLQTSFVDNVATLKILQTLDSHGQYSC 2912
QY 204 ----- 203
Db 2913 SASNPLGTASSARLTAREPKKSPFFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKD 2972
QY 204 ----- 203
Db 2973 NKEIRPGGNYITTCVGNTPHLRLIKVKGDSQVYTCQATNDVGRKMCQAQLSVKEPKFV 3032
QY 204 -----FLGTALL----- 212
Db 3033 KLEASKVAKOGESIQLECKISGSPKIVSWFRNDSSELHESWKYNMSPINSVALLTINEA 3092
QY 213 ----- 212

Db 3093 SAEDSGDYICEAHNGVGDASCSTALTAKAPPVFTQKPSVPGALKGSDVILQCEISGTPPF 3152
QY 213 ----- 212
Db 3153 EVWVKDKOVNRNKKFKITSKHFDTNLHILNLEASDVGEYHCKATNEVGSDTSCSVKPF 3212
QY 213 -----RYAQR----- 217
Db 3213 KBPPRPFVKLLSDTSTLIGDAVELRAIVEGFQPISVWMLKDRGEVIRESENTRISFDNIA 3272
QY 218 ----- 217
Db 3273 TLQGSPEASNGKYICQIKNDAGMRCSAVLTVLEPARIIEKEPEMTVTTCGNFPALECV 3332
QY 218 ----- 218
Db 3333 VTGTPELSAKWFKDGRELSADSKHHITPINKVASLKIPCAEMSDKGLYSPEVKNSVGKSN 3392
QY 219 CTHSPYL-----VNAM----- 229
Db 3393 CTVSHVSDRIVPPSFIRKLKDNVAILGASVYLECRVSGSAPISVGMFQDGNBIVSGPKC 3452
QY 230 -----SRN----- 232
Db 3453 QSSPSENCTLNLSLLEPSDTGIYTCVAANVAGSDECSAVLTVQBPSPFQTPDSVEVLP 3512
QY 233 -----LPRV----- 236
Db 3513 GMSLTPTSVIRGTPPFVKWFKGSRLELVPGESCNISSLEDFTVELELEFEVQPLESGDYSL 3572
QY 237 ----- 236
Db 3573 VTNDAGSASCTTHLFVKEPATFVKRLADFSVETGSPIVLEATYTGTPPISVSWIKDEYLI 3632
QY 237 ----- 236
Db 3633 QSERCISITWETKSTILEILESTIEDYQAQSLIENBAGQDICEALSVLEPPYFIEPLE 3692
QY 237 -----PKY----- 239
Db 3693 HVEAVIGEPAATLQCKVDGTPAIRISWYKEHTKLRSAKPAYKQPKNNVASLVINKVDHSDV 3752
QY 240 -----ING----- 242
Db 3753 GEYSCKADNSVGAVASSAVLVKARKLPFPFARKLDVHETLGPVFAPECRINGSEPIQV 3812
QY 243 ----- 242
Db 3813 SWYKGVLLKDDANLOTSFVHNVATLQILOTDQSHIGQYNCSASNPLGTASSAKLILSE 3872
QY 243 -----TKLKNT----- 248
Db 3873 HEVPPFPDLKPVSDVLDLALGESGTFKCHVTGTAPIKITWAKDNREIRPGNYKMTLVENTA 3932
QY 249 ----- 248
Db 3933 TLTVLKVGKDAGQYTCVASNIAGKSDCSAQLGVQEPFRPIKLEPSRIVKQDEFTRYEC 3992
QY 249 ----- 248
Db 3993 KIGGSPEIKVLWYKDETEIQESSKFRMSFVDSVAVLEMHNLSDSDGSDYTCEAHNAAGSA 4052
QY 249 --MEKLARKQAPVKEQFEK----- 266
Db 4053 SSSTSLVKKEPPI---FRKKPHPIETLKGAADVHLECBLOQTTPPHVSWYKDKRELRSKK 4109
QY 267 ----- 266
Db 4110 YKIMSENFLTSIHILNVDAADIGBYQKATNDVGSDDTCVGSIALKAPPRFVKKLSDISTV 4169
QY 267 -AKTQSTTT----- 275

Db 4170 VQKEVQLQTTIEGABPISVWFKDKGEIVRESDNWISYSENIATLQFSRVEPANAGKYT 4229
QY 276 ----- 275
Db 4230 CQIKNDAGMCBCFATLSVLEPATIVEKPESIKVTTGTCTLECTVAGTPELSTKWFKDGK 4289
QY 276 ----- PVP 278
Db 4290 ELTSDNKYKISFFKNVSGKLIINAPSDSGVYSPEVQNPVKDSCHTASLQVSDRTVPPSP 4349
QY 279 S----- 279
Db 4350 TRKLKETNGLSGSSVMMECKVYGPPIISVSWFHEGNEISSGRKYQTTLDNTCALTVNML 4409
QY 280 ----- TSAALNVT----- TNVTYS----- 295
Db 4410 EESDSGDTCTIATWAGSDECSAPLTVREPPSPFVQKDPDMDVLVTGNVTFTSIVKGTPPP 4469
QY 296 ----- ITTAARRVSTSTIAY- 310
Db 4470 SVSWFKGSSELVPGDRCNVSLSDSVALELFDVDTSGSGEYTCIVSNEAGKASCTHLYI 4529
QY 311 ----- 310
Db 4530 KAPAKFVKRLNDYSIEKGKPLILEGFTTGPPIISVTWKXNGINVTSPQRNITTEKSPI 4589
QY 311 ----- 310
Db 4590 LEIPSTVEDAGQVNCYNIENASGKSCSAQILILEPPYFKQLEPVKVSVDGSASLQCOL 4649
QY 311 ----- RPDSSP---WKSIMAT----- 323
Db 4650 AGTPEIGVSWYKGDTKLRPTTYTKMFRNNVATLVFNQVDINDSGEYICKAENSVEVSA 4709
QY 324 ----- QLRDL-----ATW----- 331
Db 4710 STFLTVQEQKLPSPFSRQLRDVQETVGLPVWFDCAISGSEPISVSWYKDGKPLKDSPNVQ 4769
QY 332 ----- VYTTLR----- 337
Db 4770 TSPLDNTATLNIPTKDRSLAGVSTATNPISGASSARLIITEGKNPPFDIRLAPVDA 4829
QY 338 ----- 337
Db 4830 VVGESADFECHVTGTQPIKVSNAKDSREIRSGKYQISYLENSAHLTVLKVDKGDGQYT 4889
QY 338 ----- Y 338
Db 4890 CVAVNEVGKDSCTAQLNIKERLI PPSPTKRLSETVEETEGNSFKLEGRVAGSQITVAWY 4949
QY 339 RON----- 341
Db 4950 KNNIBIOPTSNCEITFKNNTLVQVRKAGMNDAGLYTCKVNDAGSALCTSSIVIKPKK 5009
QY 342 ----- 341
Db 5010 PPVFDQHLTPVTSEGEVQLSCHVQSGSEPIRIOWLKAGRIKPSDRCSFSFASGTAVLE 5069
QY 342 ----- PFCEPSNRNRTAV----- 353
Db 5070 LRDVAKADSGDYCKASNVAGSDTTKSKVTIKDKPAVAPATKKAADVGRLLFFVSEPOSIR 5129
QY 354 ----- 353
Db 5130 VVEKTTATFIKVGGDPIPNVWTKGKWRQLNOGGRVFIHQKDEAKLEIRDTTKTDGSL 5189
QY 354 ----- 353
Db 5190 YRCVAFNEHGBIESNVNLQVDERKKQEKIEGDLFAMLKTKTPIILKKGAGEBEBIDIMELK 5249
QY 354 ----- 353
Db 5250 NVDPKVEKYARMYGITDFRGLLOAFELLKQSOBEETHRLEIBIEIERSERDEKEFEELVS 5309

QY 354 ----- 353
Db 5310 FIQORLSQTEPVTLLIKDIENQTVLKONDVAFEIDIKINYPEIKLSWYKGTLEKLEPSDKPE 5369
QY 354 ----- 353
Db 5370 ISIDGRHRTLRYKNQOLKQGNRYLRVCGPHIASAKLTVIEPAWERHLQDVTLKEGQCTWM 5429
QY 354 ----- SEPKNKTHVL----- 363
Db 5430 TVQFSPVNVKSEWFRNGRILAKPQGRHKTEVEHVKHKLTIADVRAEDQOQYTCKYEDLETS 5489
QY 364 ----- 363
Db 5490 AELRIEABPIQTKRIQNIIVVSEHSATFECEVSFDDAIIVTWYKGPTELTESESQKNFRND 5549
QY 364 ----- IRNETP-----YTIYGTI----- 376
Db 5550 GRCHWMTIHNVTDDDEGVYSVIARLEPRGEARSTAEYLTLTKEIKLELKPDPIDPSRVPI 5609
QY 377 ----- 376
Db 5610 PTMPIRAVPPEEIPPVVAPPVPLLLPTPEKKPPPKPIEVTIKKAVKDKAKKVAVAKPMBT 5669
QY 377 ----- DMSGL----- 381
Db 5670 PREBIVKPPPTTLIPAKAPEIIDVSSKABEVKIMTITRKEVQKEKEAVYEKKQAVHK 5729
QY 382 ----- 381
Db 5730 EKRVPFIESFPEPDELEVETPTPEPQPYEEDDEYBEIKVAKKEVHEWEEDEFEQG 5789
QY 382 -YY----- 383
Db 5790 EYTEREGYDEGEWEEAYQEREVIQVQKEVVEESHKRVPAKVPKAPPPPKVKKP 5849
QY 384 ----- N 384
Db 5850 VIEKIETSRMBEEKVQVTKVPEVSKKIVPQKSPRTVQEEVIEVKVPVHTKKNVISE 5909
QY 385 ETMFVENKT----- 393
Db 5910 EKMFASHTEEESVTVPEVQKEIVTEEKIHVAVSKRVEPPPKVPELPEKPAPEEVPVP 5969
QY 394 ----- 393
Db 5970 IPKKVEPPAPKVPEVKPKVPBEKKVPVPVPKPEPAAPPKVPEVPKVPPEEKIPVPVAKK 6029
QY 394 ----- 393
Db 6030 KEAPPAKVPEVQGVVTEEKIITVTORESPPPAVPEIPKKKVPBEKFPVPRKBEVPPP 6089
QY 394 ----- ASDSNKT----- 400
Db 6090 PKVPALPKVPBEKVAVPVVPAKKAPPRAEVSCKTIVVEEKRFFVAEEKLSFAPVQREVE 6149
QY 401 ----- 400
Db 6150 TRHEVSAEEMSYSBEEBEGSVISVYREBEEBEEBAEVTYEVWMBEPEYVVEEKLHIIS 6209
QY 401 ----- 400
Db 6210 KRVEAPEAVTERQEKIVLKPDKIPAKIEBPPPAKVPEAPKKIVPEKKVPAPVPPKKEKVP 6269
QY 401 ----- 400
Db 6270 PPKVPEBPKFPVEKKVPPKVIKMBEPLPAKVTEKHMQITQEEKVLNAVTKKEAPPKARV 6329
QY 401 ----- 400
Db 6330 PPEPKAVPEEKVLKLPKREBEPKAVTFRKRVVKEKVSIEAPKREBOPIKEVTIME 6389

```
QY 401 ----- 400
Db 6390 EKERAYTLTEEAASVORSEEEYBEYDYKPEEYPTTEYDQYBEYBEERYEYEEH 6449
QY 401 -----TPTSP----- 405
Db 6450 YITEPEKIPVKVPVEEPVPTKPAKPAKVLKKAIVBEKVVPVTPKPKLKPVPKVPBPK 6509
QY 406 ----- 405
Db 6510 KVFEKIHISITKREKEQVTEPAAKVPMKPKRVAAEKVPVPRKEVAPPVVRVPEVPEKE 6569
QY 406 --SMGFQRTFDPLDWYL----- 421
Db 6570 PEEVAFBEVTVHVEYLVESEESIYHBESEFITBEVVPVVPVVKVPVPRKPVEEKKP 6629
QY 422 ----- 421
Db 6630 VVPKKEAPPAKVPVEVKKEEKVPVLI PKKEKPPPAKVPVEVKVPVEKVPVVPK 6689
QY 422 ----- 421
Db 6690 VEAPPKVPVEKVPVEKVPVPAKVPKVEAPPKVPVPKLIPEEKKTPVPKKEVP 6749
QY 422 -----DSLFLDEI----- 430
Db 6750 PPKVPKREVPVPVVALPQEEVLPQEEVLPVEEVLPEEEVLPVEEVLPEEVLPEE 6809
QY 431 ----- 430
Db 6810 EEIPEEVEVPVEEYVPEEVEEVEEVLPEEVLPEEVLPEEVLPEEVLPEEVLPEE 6869
QY 431 ----- 430
Db 6870 EEAPPKVPVEKVPVEEKRIILPKKEEVLPEVTEPEEPEEPISEEEIPEEPPSIEEVEV 6929
QY 431 ----- 430
Db 6930 APPRVPEVIKAVPEAPTVPVKVEAPPKVKIPEKVPVVPVQKKEAPPKVPVEPK 6989
QY 431 ----- 430
Db 6990 VPEKKVLVPKKEAVPPAKGRTVLEBKSVAFQEVVVKERLELEVVEAEVEEIEEVEEFH 7049
QY 431 ----- 430
Db 7050 EVEEYFESEGEFHEVEEFTKLQHRVBEHREKVRHVLEVEAEVEVEFEKPKAPPKGP 7109
QY 431 ----- 430
Db 7110 ISEKIIPPKGPTKVVPKREPPAKVPVEPKIIVVEEKVRVPEEPRVPPTKVPEVLPPKEV 7169
QY 431 ----- 430
Db 7170 VPEKKVPVPPAKKEAPPKVPPEAPKVVPEKKVPVPPPKKPEVPTKVPEVPKAAVPEK 7229
QY 431 ----- 430
Db 7230 KVPEAIPPKPESPPEVEEPEEESPSAPPKKEVPPVVRVPEVKVEKVPKPAAPPK 7289
QY 431 ----- 430
Db 7290 EVTPVKVPEAPKEVVEKKVPVPPPKKEVPPTKVPEVKVAVPEKKVPEAIPPKPESP 7349
QY 431 ----- 430
Db 7350 PEVFEPEBEVALEBPAAVVEEPEAPQVTPVPKPNVPEKPAVAVVAKPELPELPVKVP 7409
QY 431 ----- 430
Db 7410 EVPKEVVEKVKVPLVVPKKEAPPKVPVEVKVEVVPKKEVAVPKPEVPPPAKVPVEPKP 7469
QY 431 ----- 430
```

```
Db 7470 VLEBKPAVFPVERAEPPPEVYBPEETAPBEEIAPBEEKVPVVAESEEPEVPPPAVPEE 7529
QY 431 -----RNFSLSRS 437
Db 7530 PKKIPEKKVPVVIKKPEAPPPEPEPEKVIKPKLKPVPVPVPVPVPVPVPVPVPVPVP 7589
QY 438 -----PTYVNLTP----- 445
Db 7590 IPKKKVPENPQVPEKVELTPLKVPGEKKVKRLKLLPERKPEKBEVVLKSVLRKRPESBP 7649
QY 446 -----PEHRAVNL----- 454
Db 7650 KVPEKKLEKVKKPAVPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBPBP 7709
QY 455 ----- 454
Db 7710 PKPEAEVKTIKPPVPEPTPIAAPVTVPVVGKAAKAPKAAKPKPIKGVPKKTPS 7769
QY 455 ----- 454
Db 7770 PIEARRKLRPGSGEKPPDEAPPTYQLKAVPLKFVKIKDIILTESEFVGSSAIFELV 7829
QY 455 STSNLSLWNLQ 465
Db 7830 SPSTAITTMWK 7840
```

RESULT 13

S20901

titin - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004

C:Accession: S20901; I46520

R:Label: S.; Gautel, M.; Lakey, A.; Trinick, J.

EMBO J. 11, 1711-1716, 1992

A:Title: Towards a molecular understanding of titin.

A:Reference number: S20897; MUID:92258380; PMID:1582406

A:Accession: S20901

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-6805 <LAB>

A:Cross-references: UNIPROT:Q28733; EMBL:X64696

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992

R:Label: S.; Barlow, D.P.; Gautel, M.; Gibson, T.; Holt, J.; Hsieh, C.L.; Francke, U.;

Nature 345, 273-276, 1990

A:Title: A regular pattern of two types of 100-residue motif in the sequence of titin.

A:Reference number: I46520; MUID:90238553; PMID:2129545

A:Accession: I46520

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 4235-5250 <LA2>

A:Cross-references: EMBL:X17329; NID:gl756; PIDN:CAA35207.1; PID:g930251

C:Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro

C:Keywords: muscle

Query Match

Best Local Similarity 19.1%; Score 469; DB 2; Length 6805;

Matches 221; Conservative 84; Mismatches 155; Indels 6096; Gaps 72;

```
QY 4 KEMMVR-DVPMFVLISIFLL----- 24
Db 52 EEVVVKEDLQKPLVDLKLKSGVLTAKGTIRLEAGVRGKPPPEVVTWKDQDFTRSR 111
QY 25 -----VS 26
Db 112 AKIDTSADSKFSLTKAKRSDGGKVVVTVTNTAGSFWAYATVNVLDKPGVRLKIPDVS 171
QY 27 FING----- 30
Db 172 SDRCTIRWDPPDDGGCEIQNYILEKCESKMWVSTYSATVLTGTTVTRLIEGNEYIFR 231
QY 31 -----KVMKALYNRP----- 41
```

Db 232 VRAENKIGTPPTESKPVIAKIDYRPGRPDPPEVTVKUSKEEMTVVMSPEYDGGKSI 291
QY 42 -----|:|:|----- 41
Db 292 YLEKKEKHSVRVFNKSAIPERRLKVNQLIPGHEYQFRVKAENEIGVGEPSLPSRPV 351
QY 42 -----|:|:|----- 41
Db 352 AKDPIPPGPINLKVDVTTKSSITLSWGRPVYDGGAPIIGYVVRPKIADASPDEGWK 411
QY 42 -----|:|:|----- 41
Db 412 RCNAAQOLVRTEFTVTSILDENOEYFRVCAQNOVGIRPAELKEAIKPEILEPPEIDLD 471
QY 42 -----|:|:|----- 50
Db 472 ASMRKLVVRAGCPIRLFAIVRGRPAKVTWRKVGIDNVKRGQVLDVDTMAFLVFNST 531
QY 51 ---GKYL-----DQLKLEIL----- 63
Db 532 RDSGKYSLTAVNPAGEKAVFNVVRVLDTPGPVSDILKVSVDYKTSCHVSWAPPENDGGSQ 591
QY 64 -----|:|:|----- 63
Db 592 VTHVIVEKRAERTWSIVNPEVKTSQVNTLVPGNEYFVRTAVNEYGPGVPADVPKP 651
QY 64 ---ROLETTISTK-----YNSKOP----- 80
Db 652 VLASDGLSEDPDPKLEVTMTKNSATLAWLPLRGGAKIDGYIISYREEDQPADRWTE 711
QY 81 ---VKNLTM-----NTEFPQ----- 92
Db 712 YSVVDLSLVITGLKEGKVKFVAARNAVGVSLPREABGVYEAKEQLIPPKILMPEQIT 771
QY 93 -----|:|:|-----YY 94
Db 772 IKAGKLRLEAHVYKGPQICWKKGEDDVTSHLAVHKAESSILIIKOVTRKDSGY 831
QY 95 ILA-----GP-----IONYSI 105
Db 832 SLTAENSSGTDQIKIVIMDRPGPPQPPFDISDIDADACSLSWHIIPLDGGSNITNIV 891
QY 106 -----|:|:|----- 105
Db 892 EKCDVSRGDVWTALASVTKTSCRIKLIPOGEYFRVRAENRFGISELPQSPKMLAQPPF 951
QY 106 -----|:|:|-----TYLWDFYSTQLR-- 118
Db 952 GVPSEPKNARVTKVANKDCIFVAWDRPDSGGSPITGYLIERKGRNSLLWVANDTAVRST 1011
QY 119 -----KPAKYYS----- 126
Db 1012 EYPCAGLEGLEYSFRIYALNKAAGSPSPKTEYVVTARTPDPGKPEVIDVTKSVSLI 1071
QY 127 -----|:|:|-----QYHNTAK 133
Db 1072 WARPKHGSGKLIIGYFVEACKLPGDKWVRCNTTHQIPHBEYTVTGLEENAQYQFRAIAK 1131
QY 134 T-----|:|:|----- 134
Db 1132 TAVNISOPSELRTPVTIHAENVPRIDLSVAMKSLLTVKAGTNVCLDATVPGKPMPTVSW 1191
QY 135 -----|:|:|----- 134
Db 1192 KKEGTVLPABGIKAMQORNICTELFVNRKOSGDYITTAENSSGSKSATIKLVLDRLP 1251
QY 135 -----|:|:|----- 140
Db 1252 GPPASVKINKMYSDRAMLSWEPPLDGGSEITNIVDKRETSRBNQAQVSNVPIITSCSV 1311
QY 141 -----|:|:|-----PCR----- 144

Db 1312 EKLEBHEYQFRICAENKYGVGDVPVFTETAIXNPDYDPPGRCDPVISNVTKDHMTYSWK 1371
QY 145 -----|:|:|----- 144
Db 1372 PPADGGSPITGYLLEKRETHAVNWKVRKPVIERTIKATGLOEGTEYEFVFTAINKAG 1431
QY 145 -----|:|:|----- 144
Db 1432 PGKPSDASKAVVAQDPLYPGPAPPKVVYDITRSSVLSWCKPAYDGGSPIGVLVEVR 1491
QY 145 -----|:|:|-----VPSMTCLSEML- 155
Db 1492 ADTNNVRCNLPOKLQKTRFEVTGLMENTEYQFRVYAVNKVGYSDPDGDKHCPKDILI 1551
QY 156 -----|:|:|-----NYSKNDTG----- 164
Db 1552 PPEGELDADLRKTLILRAGVTMRLYVPKGRPPPKITWSPKNVNLRIEGLDIKSTDFDT 1611
QY 165 -----BQCGN-----FTFNP 176
Db 1612 FLRCENVKYDAGKYILTLNESCCKGYTIWKVLDTPGPPVNVTKIEISRDSAYITWDP 1671
QY 177 -----|:|:|-----MFPNV----- 181
Db 1672 PIVDGGSPIINYVVEKDAERKSWSTVTTECSKTSFRVSNLEEGKSYFRVYAENBYGIG 1731
QY 182 -----|:|:|-----PR----- 183
Db 1732 DPGETRDAVKASETPGPVVDLKVLTWKSSCNIGWKPRSDGSRITGVVDFLTEENKW 1791
QY 184 -----|:|:|----- 183
Db 1792 QRWKSLSLOYSTKDLNEGQYTFRVSAENENGEGTSEITVVAKDDVAVDLDLKDLPD 1851
QY 184 -----|:|:|-----NTKLYV----- 190
Db 1852 LCVLAKENSFRLKI PHQKPAVSTWKKGEDPLATDTRSVSESSAVNTLVVYDCQKSD 1911
QY 191 -----|:|:|-----GPTK----- 194
Db 1912 AGKYTITLKNVAGTKBGTLSIKVGCKPGIPTGPIKPFDEVTAEBAILTKWGPDKDGSSEIT 1971
QY 195 -----|:|:|----- 194
Db 1972 NYILEKRDVNNKWTVCASAVQKTFRVTRLHEGMBEYFRVSAENKYGVGEGLKSEPIVA 2031
QY 195 -----|:|:|----- 194
Db 2032 KHPFDVPDAPPPIVVDVRHDSVSLTWDPRKTGGSPITGVHIEPKERNSLWKBEANKTP 2091
QY 195 -----|:|:|-----VNVDSTQ 201
Db 2092 IRMKDFKVTGLTEGLEEYFRVAINLAGVKPSLPSEPVVALDPIDPPGKPEVINVTRNS 2151
QY 202 IYFL-----|:|:|----- 205
Db 2152 VTLIWPYKDGKHLTGVIVEKRDPLSKTWKANHINVPDCAFTVDTLVEGKYKSEFR 2211
QY 206 -----|:|:|-----GLT----- 208
Db 2212 AKNTAGIASPSESTGTIICKDEYEAPTVILDPITKDLGLTIKAGDTIVLNAISILGKPLP 2271
QY 209 -----|:|:|-----ALLRYAOR----- 217
Db 2272 KSSWSKAGDIRPSDITQITSTPTSSMLTVKYASRKADAGEYITATNPFQTKBEHVRVT 2331
QY 218 -----|:|:|-----NC 219
Db 2332 LDVPFPGPIBISNVSAEKATLTWTPLDGGSPIKSVLEKRETSRLLTWVAEDIQSC 2391
QY 220 TH-----|:|:|----- 221
Db 2392 RHVTKLIQNEYLFRVSAVNHGKGPVQSEPVQKMDRFPFGPPGKPEVSNVTQNTAT 2451

QY 222 ----- 221
Db 2452 VSKRPTDDGSEITGYVERREKGLRWVRATKTPVSDLRCKVTGLQEGNTYFRVSAE 2511
QY 222 ----- 221
Db 2512 NRAGIGPPSDASNYVLMKDVAAPGPPSNARVTDTTKKSASLANGKPHYDGGLEITGYVW 2571
QY 222 ----- 221
Db 2572 EHQKVGDETWKDTGTPALRITEFVVPDLHTKEKNFRISAINDAGVCEPAVDPDVEIVE 2631
QY 222 -----SP 223
Db 2632 REMAPDFELDAELRRLTVRAGLSIRIFVPIKGRPAPEVTWKODINLKTRANIENTESP 2691
QY 224 ----- 223
Db 2692 TLLIPECNRYDTGKFVMTIENPAGKSGFVNVRLDTPGPVLNLRPTDITKDSVTLHMD 2751
QY 224 -----LFRV----- 236
Db 2752 LPLIDGSGRIINYIVEKREATRKSYSVTYTKCHKCTYKVTGLSEGCEYFFRVMAENEXGI 2811
QY 237 -----PKYINGTK----- 244
Db 2812 GEPSETKEPVKASEAPSPDSLINIMDITKSTVSLAWPKPKHGGSKITGYVIEAQRKGS 2871
QY 245 ----- 244
Db 2872 QMTHITTVKLECVVRNLTEGEYTFQVMVNASGRSAPRESRPVIVKBQTMPLPDLRG 2931
QY 245 -----LKNVTR----- 250
Db 2932 IYQKLVIAKAGDNIKVEIPVLRPKPTVTWKGDQVLKQTORVNVNTATSTILNISECV 2991
QY 251 ----- 250
Db 2992 RSDSGPYLTAKNIVGEVDVITIQVHDPGPPTGPIKFDEVSDFVTFWSPEPPDGGV 3051
QY 251 -----KLRKQAPVKEQPEKA----- 267
Db 3052 PISNYVIEMRQDSTTWVELATTVIRTYYKATRLTTGVEYQFRVKAQNRGVGPGITSAS 3111
QY 268 ----- 267
Db 3112 IVANYPFKVPGPPTQVTAUTKDSMTISWHEPLSDGSGPILGYHVERKERNGLMQTVS 3171
QY 268 ----- 267
Db 3172 KALVPGNIFKSSGLTDGIAEFRVIAENWACKSPKSPBVLALDPIDPPGKPIPLNIT 3231
QY 268 ----- 267
Db 3232 RHTVTLKWAPEYTGFGKITSYIVEKRLPNGRMLKANFSNILENEFTVSLGTEDAAYEP 3291
QY 268 ----- 267
Db 3292 RVIAKNAAGATSPSPSPSDAITCRDDVEAPRILVDVRFKOTVILKAGEAFKLEADVSGRP 3351
QY 268 ----- 267
Db 3352 PPTWENTKDGKELEGTGKLEIKIADPSTYLINKDSSRRSDSGAYILTATDPGPFAKHIFNV 3411
QY 268 ----- 267
Db 3412 KVLDRPGPEGLAVSEVTSEKCVLSWLPPLDDGGAKIEHYIVOKRETSRLATWNVASEV 3471
QY 268 ----- 267
Db 3472 QVTKLVTKLKGNEYIFRVMAVKNYGVGEPLSEPLVAVNPYGPDPKPNPVTITKD 3531

QY 268 -----KKT----- 270
Db 3532 SMVVCWGHPSDDGSEIINYIVERRDKAGORWVKCNKKTVDLRFKVSGLTEGHSEYEPRI 3591
QY 271 -----QSTTTPYF----- 278
Db 3592 MAENAGISAPSRTPPYKACDAVPKPGPCNPRVLDTSRSSISIAWNKPIYDGGSEITG 3651
QY 279 ----- 278
Db 3652 YMBEIALPEDEWKIVTPPPAGLKATSVITNLVENQBYKIRIYAMNSEGLGEPALVPOTP 3711
QY 279 ----- 278
Db 3712 KAEDRMLPPRIELDADLRKLVIRACCTLRFLVPIKGRPDPEVKWTRHGESLDSKASIES 3771
QY 279 --SVT-----TSAALNV----- 288
Db 3772 TSSYTLIIVGNVRPDSGKYLITVENSSGSKSAFVNRVLDTPGPPQDLKVKVTKTSVT 3831
QY 289 ----- 288
Db 3832 LTWDPPLLDGSKIKNYIVEKRESTRKAYSTVATNCHKTSWKVDQLQEGSSYFRVLAEN 3891
QY 289 ----- 288
Db 3892 EYGILPAETAESVKASERPLPPGKITLVDVTRNSVLSWEKPEHGGSRILGYIVEMQS 3951
QY 289 -----TNNVT-----YSITTA----- 300
Db 3952 KGSWKATCATVKTETATITGLQGEYSFRVSAQNEKIGISDPQLSPVPIAKOLVPPA 4011
QY 301 ----- 300
Db 4012 FKLLPNTFTVLAGEDLKIDVPPFGRPTVTWKKDDVPLKQTRVNAESTENSILLSIKE 4071
QY 301 ----- 300
Db 4072 ACREDVGHVYVYKLSNSAGEATETLNAIILDKPGPPTGPFVKMDEVTAADSIISWBPVKYD 4131
QY 301 -----RRVSTST----- 307
Db 4132 GSSNNVIVEKRDSTTTTQWISATVARTTIKASRLKTGEYQPRIAENRYKSTYLSNS 4191
QY 308 -----IAYR----- 311
Db 4192 EPIVAQYPFKVPGPPTFPVTLSSRDSMEVQWNEPVDNGGSRVIGYHLERKERNLSILWVK 4251
QY 312 ----- 311
Db 4252 LNKTPIPQTKFKTTGLEEGIEYFRVSAENIVGIGKPSKVSECYVARPCDPPGRPERII 4311
QY 312 -----PDSSFMK-----SMTATQLR----- 326
Db 4312 VTRNSVTLQWKKPYDGGSKIYGVVEKKBELPDGRWMAKSFNTIMDTQFVETGLVEDHRY 4371
QY 327 ----- 326
Db 4372 EPRVIARNAAGVSESPSESTGAITARDEIDPPRISMDPKYKDTIVVHAGESFRIDADIY 4431
QY 327 -----DLAT----- 330
Db 4432 KPIPTQWIKQDOELSNTARLEIKSTDFATSLSVKDAFRVDSGNVYLKAQNVAGERSVT 4491
QY 331 -----W-----VYTTL-- 336
Db 4492 NVKVLDRPGPEPGFIVISGVTAEKCTLAWKPLQDGGSDIINYIVERRETSRLVMTVDA 4551
QY 337 -----RY-----RONPF----- 343
Db 4552 NVQTLSCVKTKLEGNEYIFRVMAVKNYGVGEPLSEPLVAVNPYGPDPKPNPVTITKD 4611
QY 344 ----- 343

Db 4612 KDSMIVVWVWPASDGGSEILGYVLEKRDKEGIRWTRCHKELIGELRLRVTLGLENHYEF 4671
QY 344 -----CEP----- 346
Db 4672 RVSAENAGLSBPPSPAYQKACDPYKPGPPNNPKVMDITRSSVFLSKPIYDGGCEI 4731
QY 347 ----- 346
Db 4732 QGVIVEKCDVSGEWTCTPPTGINKNTNIEVEKLEKHEYNFRICAVNKAGVDHADVEG 4791
QY 347 ----- 346
Db 4792 PVIVEKLEAPDIDLLELRKIINIRAGGSLRFLVPKGRPTPEVKWGVDEIRDAAI 4851
QY 347 -----SRNR----- 350
Db 4852 DSTSFTSLVDNVNRYDSKTYTLTLENSSGTKSAFVTVRVLDTSPPPVNLKVTETKDS 4911
QY 351 -----TA 352
Db 4912 VSIWPEPLDGGSKIKNYIVEKRDSTRKSYAAVVTNCHKSSWKIDLOEGCSYFRVTA 4971
QY 353 VSEP----- 356
Db 4972 ENSYGIGLPARTADPIKVAEVPQPGKITVDDVTRNSVLSWTKPEHDGSGKIIQYIVEM 5031
QY 357 -----MKNTHVLIRN----- 366
Db 5032 QAKHSEKNECARVKSLEAVITNLTOGEEYLFRRVAVNKEGRSDPRSLAVPIVAKDLVIE 5091
QY 367 -----ETPYT-----IYGLDMSSLYY 383
Db 5092 PDVKPAFSSYVQGDLEIPEISGRPKPTITWTKOGLPLKQTRINVAOSLDLTLSI 5151
QY 384 NET-----MFVENTAS----- 395
Db 5152 KETHKDDSGHYGITVANVVGOKTASIEIITLDPDPKPGVKFDEVSASITLSNPNPLY 5211
QY 396 -----DSNKTTP----- 403
Db 5212 TGGCQITNVIYVHKRDTITVWDVVSATVARTTLKTKLTGTEYQFRIPPENRGQSAL 5271
QY 404 -----SPSMGF-----ORTFIDP 416
Db 5272 DSEPIVAQYPYKEPGPGTFFVTATSKDSMVVQWHEPINNGSPITIGYHLERKERNI-- 5329
QY 417 LWDYLDLSLLFLD----- 428
Db 5330 LMTKVDKSIIDHTQFALNLEBIEYFRVVAENIVGVKASKNSECYVARDPCDPPGTP 5389
QY 429 ----- 428
Db 5390 EAIIVKRNEITLQWTKPVYDGSMTGYIVEKRDLPGRWMAKSFNTVIEQTFTVSGLTE 5449
QY 429 ----- 428
Db 5450 DQRYEYFRVIAKNAAGTMSKPSDSTGPTAKDEVLPRIISMDPKFRDITIVNAGETFRLEA 5509
QY 429 -----EIRN-----FSLRS----- 437
Db 5510 DVHGKPLFTIEWLRGDKEVESARCEIKNTDFKALLIVKDAIRIDGGQYILRASNAVASK 5569
QY 438 -----PTTV----- 441
Db 5570 SFPVNVKVLDRPGEVPGVQVTGTCCKTLTWSPLLQDGGSDIPHYVVEKRETSRLAWT 5629
QY 442 ----- 441
Db 5630 VVASEWVNSLKITKLEGNBYIFRIMAVNKYGVGEPLSAPVLKMNPFVVPKPSLEV 5689
QY 442 ----- 441

Db 5690 TNIAKDSMTVCWNRPDSCGSGSEITGYIVEKDRSGIRWIKCNKRVRTDLRFRVTGLTETH 5749
QY 442 ----- 441
Db 5750 EYEFVRSNAAGVGEPSPATVYVKACDPVFKPGPTNAHVVDITKNSITILANGKPIYDG 5809
QY 442 ----- 441
Db 5810 GSEVLGYIIEICADDEEQIIVTPTGLKANRFBISKLIEHQYKIRVCALNKVGLGEAA 5869
QY 442 ----- 441
Db 5870 SVPGTVPKEDKLEAPELDLDELKRGIVVRAGGSARIHIPKGRPTDITWRSRGEFTD 5929
QY 442 ----- 441
Db 5930 KVQVEKGVNFTQLSIDNCNRDAGKYIVKLENSSGTKTAFVTVVKVLDTPGPPQNLAKEV 5989
QY 442 -----NLT----- 444
Db 5990 KDSAVLWPEPIIDGGAKVRNYYVIDKRESTRKAYANVSSKNKTTTFKVENLTGAIYYF 6049
QY 445 -----PPEH----- 448
Db 6050 RVMAENEFVGVVETVDVAVKAAEPPSPGKVTILTDYSQTSASLMWEKPEHDGSGRVLYG 6109
QY 449 ----- 448
Db 6110 VVENQPKTEKWSVVAESKVCNAVVTGLSSGHEYQFRVKAYNEKSGSDPRVLGVPIAKD 6169
QY 449 -----RAV----- 452
Db 6170 LTQPSKLPFKRYSVOAGEDLKIEIPVIGRPRPEIPWVKDGBPLRQTTTRVNVEETATST 6229
QY 453 -----NLT----- 456
Db 6230 ILHIKSSKDDFGKYTITATNSAGTATENLSVIVLEKPGPPVGVPRFDEISADFFVLSWE 6289
QY 457 ----- 456
Db 6290 PPAYTGCQISNVIYVEKRDTTTTWHIVSATVARTTIKTKLTGSEYQPRIYAENRYGK 6349
QY 457 -----SN 458
Db 6350 STSLDSKPVIVQYFPKPEPGPTFTVTSVSRDQMLVQWHEPVDNDGSKVLGYHLEOKEN 6409
QY 459 SLWV----- 462
Db 6410 SILWVKNKTLIQDTKFTKTLGDLGLEBYFKVSAENIVGIASLAKCPNAPVARDPCDPPG 6469
QY 463 ----- 462
Db 6470 RPEAIVITRNVTLKNWKPAYDGGSKIYIVEKKDLPGCRMWMAKSFNTVLETFVSGL 6529
QY 463 ----- 462
Db 6530 VEDQRYEYFRVIARNAAGNLSEPSSESGAITARDEIDAPNASLDPKYKDVIHVHAGETFVL 6589
QY 463 -----WLQ 465
Db 6590 EADIRKGPIDPVVWLK 6605

RESULT 14

A88852

protein unc-22 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C;Accession: A88852

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/ and
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: A8852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6831 <STO>
A:Cross-references: UNIPROT:Q23550; GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00
C:Genetics:
A:Gene: unc-22
A:Map position: 4
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 19.1%; Score 469; DB 2; Length 6831;
Best Local Similarity 3.5%; Pred. No. 2.4;
Matches 220; Conservative 92; Mismatches 131; Indels 5900; Gaps 75;

QY 2 GRKEMNR----- 9
DB 451 GOSANVKBQSDVZEELMKHKADEYQKEQSKQLQAEYKRVARRSKSKSPAP 510
QY 10 ----- 9
DB 511 QAKSTTSQGRQAEVEHKKRSSVRPDDEESQLDEIPSSGLTIPERRRELLQGVGE 570
QY 10 -----DVPKMFVLISFLLVS----- 26
DB 571 SDDEVSEGISLPSFAGGPKRKTDKPK-----KVSIAPIVSTNKSDDPESTPRRRSSI 625
QY 27 -----FINCKVMS----- 34
DB 626 DMRRESVQELKSTPLVPSGASGAPKIVEPENVTVENETAILTCVSGSPAPTR 685
QY 35 ----- 34
DB 686 WFKGSREVISGRPKHITDGEHTVALLLKCRSQDEGPYTLTIENVHGTDSADVCLAVT 745
QY 35 ----- 34
DB 746 SDNGLDFRAMLKHRESQAFQDGGGGGGGGEKKPMTBAERQSLFPQKVEKWDIP 805
QY 35 -----KALYNRP----- 41
DB 806 LPEKTVOQVDKICEWKCTYSRPNKIRWYKDRKEIFSGGLKYKIVIEKNVCTLIINPE 865
QY 42 ----- 41
DB 866 VDDTGKTCBANGVPTHAAQLTVLEPPMKYSLNPLPNTQEIYRTQAVLTCKVNTPRAPL 925
QY 42 ----- 41
DB 926 VYRGSKAIQEGDPRFIEKDAVGRCTLTIKEVEEDQAEWTARITQDVFSKVQVYVEEP 985
QY 42 -----W-----RGLV 46
DB 986 RHTFVPMKSQKVNESDLATLETVDNDKDAEYVWVHDKRIDIDGVKPKVSSNRKRLI 1045
QY 47 LS----- 48
DB 1046 INGARIEDHGEYKCTTKDRTMAQLIVDAKNKPIVALKDTEVIEKDDVTLACQTKDTP 1105
QY 49 -----KIGK-----YKLDQL----- 58
DB 1106 GIWFRNGKQISSMPGKPGFQTSRNGHTLKIGKIEADVEIDQAGLRGCNVTVLEA 1165
QY 59 -----KLEILR----- 64
DB 1166 EKRPLNWKPKIEAKAGEPCVVPFQIKGTRRGDPKQAQLKNGKPIDEMRKLVEVII 1225
QY 65 -----QLETT-----ISTKYNVSKQ- 79
DB 1226 KDDVAEIVPNPQLADTKWALELGNAGTALAPFELFVKDKPKPKGPLETK-NVTAE 1284
QY 80 -----FVK-----NLTNTEF----- 90

DB 1285 LDLVWGTPDPDEGAPVAYIIEIMQGRSGNNAKVGETKGTDFKVKDLKEHGEYKFRVKAL 1344
QY 91 ----- 90
DB 1345 NECGLSDPLTGESVLAKNPYGVPGKPKMDALDVDKDHCTLAWEPPEDGGAPITGYIE 1404
QY 91 ----- 90
DB 1405 RREKSEKDWHQGTQKPDCCCLTDKVVVEDKEYLYRVKAVNAGPGDDCDHGKPIKMAK 1464
QY 91 ---POY----- 93
DB 1465 KASPEFTGGIKDLRLKVGETIKYDVPISGBPLPECLWVNGKPLKAVGRVMSERGH 1524
QY 94 ----- 93
DB 1525 IMKIENAVRADSGKPTITLKNSSGCDSTATVTVVGRTPPKGPLDIADVCADGATLSWN 1584
QY 94 -----YILAG----- 98
DB 1585 PPDDGGDPLTGYIYEAQMDNKGKIEVGKVDPTNTTLKVNGLRNKGYKFRVKAANNE 1644
QY 99 -----PIQNY----- 103
DB 1645 GESBPLSADQYTIKDPWDEPGKPRPEITDFDADRIDIAWEPHKGGAPEIYEVVR 1704
QY 104 ----- 103
DB 1705 DPTKEWKEVVPDPTNASISGLKEGKEYQFRVAVNAGPGQSPSESEKQLAKPKPIPA 1764
QY 104 ---SIT-----YLMF----- 110
DB 1765 WLKHDNLKSIIVKAGATVRWEVKIGGEPPEVKWFKGNQOENGLQTLIDTRKNEHTILC 1824
QY 111 ----- 110
DB 1825 IPSAMRSDVGEYRLTVKNHSHGADBEKANLTVLDRPSKPNGLPVSDFEDNLNLSWKPPD 1884
QY 111 ---DFYSTQ-----LRKPAKYV----- 125
DB 1885 DDGGEPIEYVEVEKLDATGRWVPCAKVKDTKAHIDGLKGGQTYQFRVAVNKEGASDAL 1944
QY 126 -----SQY----- 128
DB 1945 STDKDKAKNPYDEPGKTGTDPVVDADRVSLWEPEPKSDGGAPIQYVIEKKKGKGRD 2004
QY 129 -----N 129
DB 2005 WQCGKVSQDQTNABILGLKEGEYQFRVAVNAGPGEASDPSRKVVAKPRNLKPWIDR 2064
QY 130 HTAKTITIFR-----PP----- 140
DB 2065 EAMKTIITIKGNDVEFPDVPVRGEPPEKKWIFNEKPVDDQKIRIESEDKYTRFVLRGATR 2124
QY 141 ----- 140
DB 2125 KHAGLYTLTATNASGDKHSVEIVILGKPSPLGPLEVSNVVEDRADLEWKVPEDDGGAP 2184
QY 141 -----PCGR----- 144
DB 2185 IDHYIEKMDLATGRWVPCGRSETTKTTPVNLQPGHEYKFRVAVNKEGESDPLTTNTAI 2244
QY 145 ---VPSMTCLEMLN----- 156
DB 2245 LAKNPYEVPGKVPKPELVUDMDKOHVLAWNAPDDGGAPIEAFVIEKKDKNGRWEALVVP 2304
QY 157 -----VSKENDTG----- 164
DB 2305 GDQKTATVPNLKEGEYQFRISARNKAGTGDPSDPSRVVAKPRNLAPRIHREDLSDTTV 2364
QY 165 -----EQCGGNFT- 172

Db 2365 KVGATLKFIHVHIDGEPADVTWSENGKIGESKAQIENEPIYSRFPALPKALRQSGKYYI 2424
QY 173 -----TEN-----PM-FFNVR 183
Db 2425 TATNINGTSDSTIINKVSKETKPKPIEVTDVDFEDRATLDMKPPEDDGGEPFIEFYBIEK 2484
QY 184 WNTK--LYV-----GPTKVNVDS-----199
Db 2485 MNTKGIWVPCGRSGDTHFTVDSLNGDHYKFRKAVNSGSPDPLETETDILAKPPFR 2544
QY 200 -----QTIYF 204
Db 2545 PDRGRPEPTDWDSDHVDLKWDPPLSDGAPIEBEYQIEKTKYGRWEPALTVPGGT---2601
QY 205 LGLTA-----209
Db 2602 --TATVPDLTPNEEYFRVAVNVKGGSPDSASKAVIAKPNLKHIDRDALKNLTK 2658
QY 210 -----LLRYAQRNCTHSF- 223
Db 2659 AGOSISFDPVPSGEPAPTWTWHWPDNREIRNGGRVKLDNPEYQSKLVVKQMERGDSGTFT 2718
QY 224 -----223
Db 2719 IKAVNGEDEATVKINVIDKPTSPNGPLDVSDVHGHTVLNWRAPDDGGPIENYVIE 2778
QY 224 -----223
Db 2779 KYDTASGRWVPAKVAGDKTTAVVDGLIPCHEYKFRVAANAEGSDPLETFTGLAKDP 2838
QY 224 -----223
Db 2839 FDKPGKTNAPEITDWDKHVDLEWKPPANDGGAPIEBEYVEMKDBSPFWNDVAHVPAQ 2898
QY 224 -----223
Db 2899 TNATVGNLKEGSKYEFIRAKNKAGLDPSDSASAVAKARNVPPVIDRNSQIEIKVKAGQ 2958
QY 224 -----YLVNAM 229
Db 2959 DFLSNIPVSGEPTTITWTFTGTPVESDDRMKLNEDGKTKPHVKRLRSDTGTIYIKAE 3018
QY 230 SRN-----232
Db 3019 NENGTDTAEVKTVLDHPSSPRGLDVTNIVKDCDLAWKEPDDGGAEISHYVIEKQDA 3078
QY 233 -----232
Db 3079 ATGRWTACGESKDTNFHVDLLTQGHYKFRKAVNRHGDSPLEAREALIAKDPDRADK 3138
QY 233 -----232
Db 3139 PGTPEIVDWDKHADLKWTPPADGGAPIEGYLVEMRTPSGDWVPAVTVGAGELTATVDG 3198
QY 233 -----LFRV-----PKVI-----240
Db 3199 LKPGQTYQFRVKALNKAGESTPSRFTWAKPHLAPKINRDMFVAQRVAGQTLNFDV 3258
QY 241 -----NKTML--NKTMR-----250
Db 3259 NVEGEPAPKIEWFLNGSPSSGGNTHIDNNTDNNKLTSTKSTARADSGKYIVATNESGK 3318
QY 251 -----250
Db 3319 DEHEVDVNILIPGAPGFLRHKOITKESVVLKWDEPLDDGGSPITVYVVEKQDGRWV 3378
QY 251 -----KLK-----253
Db 3379 PCGETSDTSLKVNKLSEGHYKFRKAVNRQTSAPLTSDBAIVAKNPPDPDPTDTP 3438
QY 254 -----253
Db 3439 VDWKDHVDLEWKPPANDGGAPIDAYIVEKKDKFGDWVECARVDGKTKATADNLTPGET 3498

QY 254 -----RKOAP-----258
Db 3499 YQFRVKAVNKAGPGKPSDPTGNVAKPRMAPKLNLAGLLDIRIKAGTPIKLDIAFEGEP 3558
QY 259 -----258
Db 3559 AFVAKWANDATIDTGARADVNTPTSSAIHIFSAVRGDTGVYKLIIVENEHKGDTAQCNV 3618
QY 259 -----258
Db 3619 TVLDVPGTPEGFLKIDEIHKEGCTLNWKPTDNGTDLVHYIVEKMDTSGRTGWEVGTFF 3678
QY 259 -----VKEQF-----263
Db 3679 DCTAKVNKLVPKVEYAFRVKAVNLQGESKPLEAESEPIIAKNQFDVDPDPKPEVTDWDKD 3738
QY 264 -----EKKAKK---TOSTTTPYFSYTTSAALNVTTNVTY---294
Db 3739 RIDIKWNPNTANNGAPVTGYIVEKKEGSAIWTGKTP--GTTFSADNLKPGVEYEF 3795
QY 295 -----SITTAARRV-----303
Db 3796 VIAVNAAGSPSPDPTDPQITKARYLKPILTSRKKIKAGFTHNLEVDFIGAPDPTAT 3855
QY 304 -----STSTIAY-----310
Db 3856 WTVGDSGAALAPELLVDAKSTTSIFPPSAKRADSGNYKLKVNKELGEDEAIFEVIVQDR 3915
QY 311 -----310
Db 3916 PSAPEGPLEVSDVTKSCVLNWKPPKDDGGAESNYVVEKRDTKTNTWVPSAFVTGTSI 3975
QY 311 -----310
Db 3976 TVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFGTGPCKGPRPEIVTDNDHIDI 4035
QY 311 -----310
Db 4036 KNDPPRDNGGSPVDHYDIERKDAKTRWIKVNTSPVQGTAFSDTRVQKGHTYEYRVAVN 4095
QY 311 -----RPDSSEPMK 318
Db 4096 KAGPGQPSDSSAAATAKPMHEAPKFDLDLDGKEFRKAGEPLVITTFPTASQPDLSWK 4155
QY 319 S-----319
Db 4156 EGGKPLAGVETTDTSOTKLVIPSTRSDSGPVKIKAVNPYGEAEANIKITVIDKPGAPENI 4215
QY 320 -----319
Db 4216 TYPVSRHTCTLNWDAPKDDGGAETAGYKIEYQEVGSQIWDKVPGLISGTAYTVRGLBHG 4275
QY 320 -----319
Db 4276 QOYFRFIRAENAVGLSDYCCQGVVVKDPDPDPGAPSTPEITGYDTNQVSLAWNPRDDG 4335
QY 320 -----IMATO-----324
Db 4336 GSPILGYVVERFEKRGGDWAPVKMPWKGTCEIVPGLHENETYQFRVAVNAAGHGEPS 4395
QY 325 -----324
Db 4396 NGSEPVTCRPVVEKPGAPDAPRVGKITKNSAELTWNRLRDGGAPIDGYIVEKKKLGDN 4455
QY 325 -----LRDLA-----329
Db 4456 WTRCNDKVPDRTAFEVKNLGEKEEYFRVIAVNSAGEGEPSPDLVLIEBQGRPFDI 4515
QY 330 -----329
Db 4516 NNLKDIITVRAGETIQIRIPYAGGNPKPIIDLFGNSPIFENERIVVDNPGEIUITTGS 4575

QY 330 -----TW----- 331
Db 4576 KRSAGPYKISATNKYKDKTCLNVFLDAPGKPTGPIRATDIOADAMTLSWRPPKONGG 4635
QY 332 -----VYTLRYR----- 339
Db 4636 DAIYVYVEKRTPGGDWTVGHPVGTTLRVRNLDANTPYEPRVRAENQYGVGEPLTDDA 4695
QY 340 -----QNPFCPEP----- 346
Db 4696 IVAKNPFDTGAPGQPEAVETSEBALTQWTRPTSDGAPIQGVYIEKREVGSTEWTKAA 4755
QY 347 -----SENRTA----- 352
Db 4756 FGNILDTKHRVTGLTPKTYEPRVAAYNAAGGEYSVNSVPITADNAPTRPKINNGMLTR 4815
QY 353 ----- 352
Db 4816 DILAYAGERAKILVPPAASPAPKVTFSKGENKISPTDPRVKVEYSDPLATITIEKSELTD 4875
QY 353 ----- 352
Db 4876 GGLYFVELENGSGDSASIRLKWVDKPPASQHIRVEDIAPDCCTLYWMPSSDGSPTN 4935
QY 353 -----VSEPMKTHVL-----IRNETPYTI----- 372
Db 4936 YIVEKLDLRHSDGKWEKVSFVRNLNTYVGGLIKDNRVFRVRAETQYGVSEPCELADW 4995
QY 373 ----- 372
Db 4996 VAKYQFVNPQPEAPTVRDKDSTWAELEWDPDRGGSKIICYQVQYRDTSSGRINAKMD 5055
QY 373 ----- 372
Db 5056 LSEQCHARVTLRQNGEPEPRIIAKNAAGSPKSPSPSERCQLKSRFGPPGPIHVGAKSI 5115
QY 373 -----YG----- 374
Db 5116 GRNHCTITWMAPLEDGGSKITGYNVEIREYGSTLTWASDVNVRBEPTVDKREFNDYE 5175
QY 375 ----- 374
Db 5176 PRVAINAGKIFSLPSGPIKIQESGSRPQIVVKPEDTAQPNRRRAVTFCEAVGRPEP 5235
QY 375 -----TLDMSLLYNETMFV----- 389
Db 5236 TARWLNRGRELPESSRYRFEASDGYKFTIKEVWDIDAGEYTVESNPNYGSDTATANLV 5295
QY 390 ----- 389
Db 5296 QAPVIEKDVNTILPSGDLVRLKIYFSGTAPFRHSLVLRNEEDMDHPTIRIVEFDDHI 5355
QY 390 -----ENKTASDSNKT----- 401
Db 5356 LITIPALSVREAGRYEYTVSNDSEATGFWLNTGLPEAPQGPLHISNIGPSTATLSWR 5415
QY 402 ----- 401
Db 5416 PPVTDGSKITSYVVEKRDLSKDEWTVTSNVKDMNYIVTGLFENHEYEPRVSAQENGI 5475
QY 402 -----PTSP-----SMGFORTFIDPLWD----- 419
Db 5476 GAPLVSEHPHIIARLPFPDPPTSPLEIVQVGGDVYTLWSQR-----PLSDGGGRLRGYIVE 5531
QY 420 -----YLDLSLF----- 426
Db 5532 KQEBEHDWFRCNQPNPPNNYVNLIDGRKYRYRVPVANDAGLSDLAELDQTLFOASG 5591
QY 427 ----- 426
Db 5592 SGECPKIVSPLSULNEEVGRCVTFECEISGSPREPRWFKGCKELVDTSKYTLINKGDKQ 5651
QY 427 ----- 426

Db 5652 VLIINDLTSDDADEYTCRATNSSGSTRANLRIKTKPRVFIPPKYHGGYEAQKGETIEL 5711
QY 427 ----- 426
Db 5712 KIPYKAYPQGEARWTKDGEKIENNSKFSITTTDDKFATLIRISNASREDYGEYRVVVVSG 5771
QY 427 -----LDEI----- 430
Db 5772 SDSGTVNVTVADVPEPPFPPIIENILDEAVILSWKPPALDGGSLVTNVTIEKREAMGGSW 5831
QY 431 ----- 430
Db 5832 SPCAKSRYTYTISGLRAGKQYEFRIIAENKHGSKCPCPTAPVLIPODERKRRRGYDVD 5891
QY 431 ----- 430
Db 5892 EQGKIVRGKTVSSNYVFDIWKQYYPQVEIKHDHVLHDYDIHEELGTGAFGVVHRV 5951
QY 431 -----RNFSS-----LRSPYVNL----- 443
Db 5952 TERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAFEDDNEVMYIYBFM 6011
QY 444 ----- 443
Db 6012 SGGELFEKVADEHNKMSDEBAVEYMRQVCKGLCHMHNHYVHLDLKPENIMPTTKRSNEL 6071
QY 444 ----- 443
Db 6072 KLIDPGLTAHLDPKQSVKVTGTAEFAAPEVAEKPVGYYTDMHSGVLSVILLSGLSPP 6131
QY 444 ----- 443
Db 6132 GGENDETLRNVKSCDMMDSDAFSGISEDGKPIRKLLADPNTRMTIHOALEHPWLTP 6191
QY 444 -----TPP-----BHR----- 449
Db 6192 GNAPGRDSQIPSSRYTKIRDSIKTKYDAMPPLPGLGRISYSLRKHQRPQEIIRDAFW 6251
QY 450 ----- 449
Db 6252 DRSEAQRPIVKPYGTEVGEGQSANFYCRVIASSPPVVVTHKDDRELKQSVKYMRYNGN 6311
QY 450 -----RAVN----- 453
Db 6312 DYGLTINRVKDDKGVTYVRAKNSYGTKEBIVFLNVTNRHSBPLKPEPLEMKKAPSPRV 6371
QY 454 ----- 453
Db 6372 BEFKERSAPFTFHLNRNLIQKNHQCKLTCSLQGNPNPTIEMKDGHPVDEDRVQVSFR 6431
QY 454 ----- 453
Db 6432 SGVCSLEIFNARVDDAGTYTATNDLGDVSECVLTVQTKGGEPIPRVSSFRPRRAYDT 6491
QY 454 LST----- 456
Db 6492 LSTGTOVERSHSYADMRRRSLIRDVSPDVSRAADDLTKITNELPSTFQAQLSDSETEVG 6551
QY 457 ----- 456
Db 6552 SAEFSAAVSQGEPLIEWLHNGERISESDSRFRASYVAGKATLRISDAKKSDEGQYLCRA 6611
QY 457 SNS----- 459
Db 6612 SNSAGQEQTRATLTVKGDQPLNQHAGQAVESELRVTKHLGGEIVNNGESVTFEARVQGT 6671
QY 460 -----LM----- 461
Db 6672 PEEVLWNRNGELTNGDKTISIQDGETLSFTINSADASDAGHYQLEVRKGTNLVSVASL 6731
QY 462 -----WW 463
|:

Db 6732 VVGEKADPPVTRLPSSVSAPLGGSTAFTEPENEVLTVQWF 6774

RESULT 15

S57242

twitchin [similarity] - Caenorhabditis elegans

N:Alternate names: myosin-regulating protein

N:Contains: protein kinase (EC 2.7.1.1-)

C:Species: Caenorhabditis elegans

C:Date: 28-Oct-1995 #sequence revision 24-Oct-1997 #text change 09-Jul-2004

C:Accession: S57242; S07571; S06797; S57218; T27934; T28030

R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.

submitted to the EMBL Data Library, February 1993

A:Description: Additional sequence complexity within twitching of Caenorhabditis elegans

A:Reference number: S57242

A:Accession: S57242

A:Molecule type: DNA

A:Residues: 1-6839 <BEN1>

A:Cross-references: UNIPROT:Q23550; EMBL:L10351

A:Experimental source: var. Bristol

R:Benian, G.

submitted to the EMBL Data Library, November 1989

A:Reference number: S07571

A:Accession: S07571

A:Molecule type: DNA

A:Residues: 792-6839 <BEN2>

A:Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898

A:Experimental source: var. Bristol

R:Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.

Nature 342, 45-50, 1989

A:Title: Sequence of an unusually large protein implicated in regulation of myosin activity

A:Reference number: S06797; MUID:90044042; PMID:2812002

A:Accession: S06797

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 806-1175; 1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-5693; 5696-6359, 'I', 6360-6839

A:Cross-references: EMBL:X15423

A:Experimental source: var. Bristol

R:Benian, G.M.; L'Hernault, S.W.; Morris, M.E.

Genetics 134, 1097-1104, 1993

A:Title: Additional sequence complexity in the muscle gene, unc-22, and its encoded protein

A:Reference number: S57218; MUID:93387664; PMID:8397135

A:Accession: S57218

A:Molecule type: DNA

A:Residues: 2-99; 108-194, 'Q', 196-206; 374-468; 658-753 <BEN4>

A:Experimental source: var. Bristol

R:White, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20442

A:Accession: T27934

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 'MGIPGKKCKQ', 19-6839 <W1>

A:Cross-references: EMBL:Z73897; PIDN:CAA98064.1; GSPDB:GN00022; CESP:ZK617.1a

A:Experimental source: clone ZK617

R:Harris, B.

submitted to the EMBL Data Library, May 1996

A:Reference number: Z20458

A:Accession: T28030

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 'MGIPGKKCKQ', 19-6839 <W2>

A:Cross-references: EMBL:Z73899; PIDN:CAA98081.1; GSPDB:GN00022; CESP:ZK617.1a

A:Experimental source: clone ZK829

C:Comment: Lack of unc-22 leads to a constant twitching of the body muscles.

C:Genetics:

A:Gene: unc-22; CESP:ZK617.1a

A:Map position: 4

A:Introns: 18/3; 69/3; 143/2; 176/3; 264/2; 387/3; 413/2; 471/1; 516/3; 550/3; 582/3; 601/3; 691/3; 6776/1; 6808/3

C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

C:Keywords: ATP; autophosphorylation; duplication; muscle; phosphotransferase; serine/threonine kinase

F:806-898, 899-990, 991-1083, 1084-1175, 1178-1273, 1474-1567, 1770-1864, 2066-2158, 2358-2450, 2451-2452

96-5790,6263-6356,6386-6478,6541-6635,6649-6742,6745-6838/Region: motif 2

F:1274-1372,1373-1473,1568-1670,1671-1769,1865-1964,1965-2065,2159-2258,2259-2357,2451-2452

F:5940-6197/Domain: protein kinase homology <KIN>

F:5948-5956/Region: protein kinase ATP-binding motif

F:5971/Active site: Lys #status predicted

Query Match 19.1%; Score 469; DB 2; Length 6839;

Best Local Similarity 3.5%; Pred.No.2.4;

Matches 220; Conservative 92; Mismatches 131; Indels 5900; Gaps 75;

QY 2 GRKMMVR----- 9

Db 459 QSSAMVRIEQSDVEELMKHDKDAEDYQKEQKSQLQAETFKRVARRSKSKSPAP 518

QY 10 ----- 9

Db 519 QAKSTTSSEGRQAESEVHKRSSVRDPDEESQLDEIPSSGLTIPEERRRLLGQVGE 578

QY 10 -----DVPKMFVLISIFLLVS----- 26

Db 579 SDDEVSEISELPSFAGGKPRRKTDKPK-----KVSIAFVSTNKSDDDPSTPRRSSI 633

QY 27 -----FINCKVMS----- 34

Db 634 DMRRESVQIELEKTSFPLVPSGASGAPKIVEPVENVVNETALTCKVSGSPAPTR 693

QY 35 ----- 34

Db 694 WFKGSREVISGGRFKHITDKEHTVALALLKCRSQDEGYTLTIENVHGTDSADVLLVT 753

QY 35 ----- 34

Db 754 SDNGLDFRAMLKHRSQAGFQKDGEGGAGGGGKKPMTAEARRQSLPFGKKVKEWDIP 813

QY 35 -----KALYNRP----- 41

Db 814 LPEKTVQQQVDKICEWKCTYSRPNKIRWKORKRIFSGGLKYKIVIEKNVCTLIINPE 873

QY 42 ----- 41

Db 874 VDDTGKTCANGVPHTAQLTVLEPPMKYSFLNPLPNTQEIYRTKQAVLTCKVTPRAPL 933

QY 42 ----- 41

Db 934 VMYRGSKAIOEGDPRFIIEKDAVGRCTLTIKEVEEDDQAEWTARITQDVFSKVQVYVEP 993

QY 42 -----W-----RGLV 46

Db 994 RHTFVPMKSQKVNESDLATLETVDYDKDAEVDVWHDKRIDIGVKFKVSSNRKRLLI 1053

QY 47 LS----- 48

Db 1054 INGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFVALKDTVEIKDDVTLTCQTKDKTP 1113

QY 49 -----KIGK-----YKLDOL----- 58

Db 1114 GIWFRNGKQISSMPGKPEFQSRNGHTHLKLGKIEKMEADVYETDQAGLRGSCNVTVLEA 1173

QY 59 -----KLEIR----- 64

Db 1174 EKRPLNWKPKIEBAKAGEPCVVKVPFQIKGTRRGDPKAILKNGKPIDEMRKLVEVII 1233

QY 65 -----QLETT-----ISTKYNVSKQ- 79

Db 1234 KDDVAEIVFKNPQLADTGKVALELGNAGTALAPFELFVKDKPKPKPGPLETK-NVTAG 1292

QY 80 -----PVK-----NLTMTEF----- 90

Db 1293 LDLYVMGTDFDEGAPVKVAYIIEMQEGRGNWAKVGTEKGTDFVKVLDKHEGVEKPRVKAL 1352

QY 91 ----- 90

Db 1353 NECGLSDPLTGESVLAKNPYGVPGKPKMMDAIDVDKDHCTLAWBPPBDDGGAPITGYIIE 1412
QY 91 ----- 90
Db 1413 RREKSEKDWQGVQTKPDCCELTDKKVVEDKEYLYRVKAVNKAGPGDPCDHGKPIKMKAK 1472
QY 91 ---POY----- 93
Db 1473 KASPEFTGGIKDLRLKLVGETIKYDVPISGEPLECLWVVGKPLKAVGRVMSSEBGRKH 1532
QY 94 ----- 93
Db 1533 IMKIENAVRADSGKFTITILKNSSGSCDSTATVTVVGRPTPPKGPLDIADVCADGATLSWN 1592
QY 94 -----YILAG----- 98
Db 1593 PPDDGGDPLTGYIVEAQMDNKGKYEIVGVKVDPNNTTLKVNGLRNKGNYKFRVKAVNNE 1652
QY 99 -----PIQNY----- 103
Db 1653 GESEPLSADQYTOIKPDWDEFGKGRPEITDFDADRIDIAWEPHKGAGAPIEIVEVR 1712
QY 104 ----- 103
Db 1713 DPDTKEWEKVRVDPDTNASISGLKEGKEYQFRVRAVNAKAGPGQSPSEKQOLAKPKFIPA 1772
QY 104 -----SIT-----YLF----- 110
Db 1773 WLKHDNLKSIIVKAGATVRWEKIGBEPPEVKWPKGNQLENGIQLTIDTRKNEHTILC 1832
QY 111 ----- 110
Db 1833 IPSAMRSDVGBYRLTVKNSHGADEKANKLTVDLRPSKPNGLPVSVDVPEDNLNLWSKPPD 1892
QY 111 -----DFYSTQ-----LRKPKYVY----- 125
Db 1893 DGGEPPIEYVEKELDTATGRWVPCAKVKOTKAHIDGLKGGTQYQFRVKAVNKAGSADAL 1952
QY 126 -----SOY----- 128
Db 1953 STDKDTAKNPYDEPGKTGTPDVVDMDADRVSLEWEPKSDGGAPITQYVIEKKHGRD 2012
QY 129 -----N 129
Db 2013 WQEGKVSQDQTNABEILGLKEGEYQFRVKAVNAKAGPGASDPSRKVAVKPNLKPWIDR 2072
QY 130 HTAKTITFR-----PP----- 140
Db 2073 EAMKTITIKVGNVDFVVPVGEPPPKKEWIFNEKPVDDQKIRIESEDYKTRFVLRGATR 2132
QY 141 ----- 140
Db 2133 KHAGLYTLTATNASGDKHSVEVIVLKGPSPLGLEVSUVYEDRADLEWKVPEDDGGAP 2192
QY 141 -----PCGR----- 144
Db 2193 IDHVEIEKMDLATGRWVPCGRSETTKTVPNLQPGHEYKFRVRAVNAKAGESDPLTTNTAI 2252
QY 145 -----VPSMTCUSEMLN----- 156
Db 2253 LAKNPYEVGKVDKPELVWDKOHVDLAWNAPDDGGAPIEAFVIEKKDKNGRWEALVVP 2312
QY 157 -----VSKRNDTG----- 164
Db 2313 GDQKTATVNLKEGEYQFRISARNKAGTDPSPDRVAKPNLAPRIHREDLSDTTV 2372
QY 165 -----EQCGNFT----- 172
Db 2373 KVGATLKFIVHIDGEPAPDVTWSFNGKIGESKAIEENEPYISRFALPKLRKQSGKTYI 2432
QY 173 -----TFN-----PM-FFNVPR 183
Db 2433 TATNINGDSVTINIIVKSKTPKGPPIEVTVDFEDRATLDWKPPEDDGGEPPIEYIEK 2492

QY 184 WNTK--LYV-----GPTKVNVD----- 199
Db 2493 MNTKOGIWWPCGRSGDTHFTVDSLNGKDHKFRKAVNSEGSDPLETETDILAKNPPDR 2552
QY 200 -----QTYLF 204
Db 2553 PDRGRPEPTDWDSDHVDLKWDPPLSDGGAPIBIBYQIEKRTYGRWEPAITVPGGQT--- 2609
QY 205 LGLTA----- 209
Db 2610 ---TATVPDLTPNEBYEYFVRVAVNKGSPSDASKAVIAKPRNLKPHIDRDALKNLTIK 2666
QY 210 -----LLRYAQRNCTHGF- 223
Db 2667 AGQISIFDVPVSGSPAPTIVTWHWPNRREIRNGRVRKLDNPEYQSKLVVKQMERGDSGTFT 2726
QY 224 ----- 223
Db 2727 IKAVNANGEDBATVKINVIDKPTSPNGPLDVSVDHGHVTLNWRAPDDDDGIPNIYVIE 2786
QY 224 ----- 223
Db 2787 KYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAAVNAEGESDPLETFTGTTLAKDP 2846
QY 224 ----- 223
Db 2847 FDKPKTNAPEITDWDKDHVDLEWKPPANDGGAPIEYVYVEMKDEFPFPMNDVAHVPAQ 2906
QY 224 ----- 223
Db 2907 TNATVGNLKEGSKYEFIRAKNKNAGLGDPSDSASAVAKARNVPPVIDRNSIOEIKVKAGQ 2966
QY 224 -----YLVNAM 229
Db 2967 DFLNIPVSGEPTTITWTFEGTPVESDDRMKLNEDGKTKFHVKRALLRSDTGYIIEK 3026
QY 230 SRN----- 232
Db 3027 NENGTDTAEVKVTVLDHPSSPRGLDVTNIVKDCDLAWKEPEDDGGAEISHYVIEKQDA 3086
QY 233 ----- 232
Db 3087 ATGRWTAGESKOTNFHVDLDTQGHYKFRVKAVNRHGDSDPLEAREAIIAKDPDRADK 3146
QY 233 ----- 232
Db 3147 PGTPEIVDWDKDHADLKWTPPADGGAPIEGYLVEMRTPSGDWVPAVTVGAGELTATVDG 3206
QY 233 -----LFRV-----PKYI----- 240
Db 3207 LKPGQTYQFRVKALNKAGESTPSPSRMTWAKPRHLAPKINRDMFVQVRKAGQTLNFDV 3266
QY 241 -----NGTKL--KVTMR----- 250
Db 3267 NVEGEPAPKIEWFLNGSLSSGGNTHIDNNNTDNNTKLTTSSTARADSGKYKIVATNESGK 3326
QY 251 ----- 250
Db 3327 DEHEVDVNIIDIPCAPSGPLRHKDITKESVVLKWDPELDGSGPITVYVVEKQEDGGRWV 3386
QY 251 -----KLK----- 253
Db 3387 PCGETSDSLKVNKLSEGHEYKFRVKAVNRQGTSAPLTSDHAIIVAKNPFDEPDAPTDVTP 3446
QY 254 ----- 253
Db 3447 VDMKDHDVLEWKPANDGGAPIDAYIVEKKDFGWVECARVDGKTTKATADNLTPGET 3506
QY 254 -----RKQAP----- 258
Db 3507 YQFRVKAVNAKAGPKSPDPTGNVVAKPRRMAPKUNLAGLLDLRIKAGTPIKLDIAFEGEP 3566

QY 259 ----- 258
Db 3567 APVAKWKANDATIDTGARADVNTPTSSAIHIFSAVRGDTGYKIIVENEHGKOTACNV 3626
QY 259 ----- 258
Db 3627 TVLDVFTPBGLIKIDEIHKEGCTLNWKPTDNGGTDVLHYIVEKMDISRGTWQEVGTFP 3686
QY 259 ----- VKQF ----- 263
Db 3687 DCTAKYNKLVPGKEYAFRVAKNVLOGESKPLAEBPIIAKNQFVDPDPVKPEVTDMDKD 3746
QY 264 ----- EKKAKK ----- TQSTTTTPYFSYTTSSAALANVTWY ----- 294
Db 3747 RIDIKWNPANNGAPVTGYIVEKKEKGSALTWEAGTPT --- GTTFADNLKPGVGEYFR 3803
QY 295 ----- SITTAARRV ----- 303
Db 3804 VIAVNAAGSDSPDPTDPOITKARYLKPKILITASKIKIRKAGFTHNLEVPFIGAPDPTAT 3863
QY 304 ----- STSTIAY ----- 310
Db 3864 WTVGDSGAALAPELLIVDAKSTTSIFPFSAKRADSGNYKLVKNELGEDEAIFEIVQDR 3923
QY 311 ----- 310
Db 3924 PSAPGPLEVSDVTKDSVNLNWKPPKDDGAEISNYVVEKRDTKTNTWVPVSAFVTGSI 3983
QY 311 ----- 310
Db 3984 TVPKLTGHEYEYFRVMAENTFGRSDSLNTEPVLAKDPFGTGPKGORPEIVTDNDHIDI 4043
QY 311 ----- 310
Db 4044 KWDPRNGSPVDHYDIERDKAKTGRWIKVNTSPVQGTAFSDTRVQKHGHTYEYRVAVN 4103
QY 311 ----- RPSGFMK 318
Db 4104 KAGPGQPSDSSAAATAKPMHEAPKFDLDLGKEFRVKAGEPLVITPFTASPQPDISWTK 4163
QY 319 S ----- 319
Db 4164 EGKPLAGVETSDQTLVIPSTRSDSGPVKIKAVNPYGEAEANIKITVIDKCAPENI 4223
QY 320 ----- 319
Db 4224 TYPAVSRHTCTLNWDAPKDDGAEAGYKIEYQEVGSQIWDKVPGLISGTAYTVRGLHG 4283
QY 320 ----- 319
Db 4284 QOYRFRIRABNAVGLSDYCGVPVVIKOPFDPGAPSTPEITGYDTNQVSLAWNPPRDDG 4343
QY 320 ----- IWATO ----- 324
Db 4344 GSPILGVYVERFEKRGGDWAPVMPVMVKGTECIVPLGHENETYQFRVRAVNAAGHGEPS 4403
QY 325 ----- 324
Db 4404 NGSEPVTCRYYVEKPGAPDAPRVGKITKNSAELTNRPDRDGGAPIDGYIVEKKLGDND 4463
QY 325 ----- LRDLA ----- 329
Db 4464 WTRCNDKPRDTAFEVNKLGEKEYEYFRVIAVNSAGEBPSKPSDLVLIEQPGRPIFDI 4523
QY 330 ----- 329
Db 4524 NNLKDI TVRAGETIQIRIPVAGGNPKPIIDLFNNGSPIFENERTVVDVNPGEIVITTTGS 4583
QY 330 ----- TW ----- 331
Db 4584 KRSDAGPYKISATNKYKDKCKLVNFVLDAPGKPTGPIRATDIOADAMTLSWRPPKONGG 4643
QY 332 ----- VYTLTRYR ----- 339

Db 4644 DAITNTVVSKRTPGGDWVTGHPGVGTTLRVRLNDANTPVEFRVRAENQYGVGEPLETDDA 4703
QY 340 --- QNPFCBP ----- 346
Db 4704 IVAKNPFDTPGAPGQPEAVEISEAITLQWTRPTSDGGAPIQGVVIEKREVGSTEWTKAA 4763
QY 347 --- SRNRTA --- 352
Db 4764 FGNILDTKHRVTGLTPPKTYEYFRVAAYNAAGQGEYSVNSVPIITADNAPTRPKINWMLTR 4823
QY 353 --- 352
Db 4824 DILAVAGERAKILVPPAASPAPKVTFPSKGENKISPTDPRVKVEYSDFLATLTIKSELTD 4883
QY 353 --- 352
Db 4884 GGLYFVELENSQSGDSASIRLKVVDPKASPQHIVERVEDIAPDCCTLYWMPSPSSGGSPITN 4943
QY 353 --- VSEFMKNTHVL --- IRNETPTYI --- 372
Db 4944 YIVEKLDLRHSDGKWEKVSSFVRNLZNTVGGLIKONRYRFRVRAETQYGVSEPCELADV 5003
QY 373 --- YG --- 374
Db 5124 GRNHCTITWMAPLEDGGSKITGVNVEIREYGSTLTWASDYNVREPEFTVDKLRFNDEY 5183
QY 375 --- 374
Db 5184 FRVVAINAAGKIPSLPSGPIKIOBSSGSRPOIVVVKBEDTAQYNNRAVFTCEAVGREP 5243
QY 375 --- TLDMSLYNETMFV --- 389
Db 5244 TARWLENGRELPESSRYRFEASGVYKFTIKEVWMDIDAGEYTVESVNFYSDTATANLV 5303
QY 390 --- 389
Db 5304 QAPPVIEKDVNTIILPSGDLVRLKIYFSGTAPFRHSLVLRNREIDMDHPTIRIVEFDDHI 5363
QY 390 --- ENKTASDSNKT --- 401
Db 5364 LITIPALSVREAGRYEYTVSNDSGEATTFWLVNTGLPEAPOGLHISNIGPSTATLSWR 5423
QY 402 --- 401
Db 5424 PVTGGSKITSYVVEKRDLSKDEWTVTSNVKDMNYIVTGLFENHEYEYFRVSAQNGEI 5483
QY 402 --- PTSP --- SMGFORTFIDPLWD --- 419
Db 5484 GAPLVSEHPILARLPDPPTSPNLLEIVQYGGDYVTLISWOR --- PLSDGGGRLRGYIVE 5539
QY 420 --- YLDSLLF --- 426
Db 5540 QOBEHDEWFRCONPSPNNYNVPLNIDGRKYRYRFAVNDAGLSDLAELDQTLFQASG 5599
QY 427 --- 426
Db 5600 SGEPKIVSPLSLDNEBVGRCVTFCEISGSRPPEYRWFPGCKELVDTSKYTLINKDKQ 5659
QY 427 --- 426
Db 5660 VLINDLTSDDADEYTCRATNSGSTRANLRITKPRVFIPIPKYVHGVEAQKGETIEL 5719
QY 427 --- 426

Db 5720 KIPYKAYPGGEARWTKDGEKIENNSKFSITTTDDKFA TLRI SNASREDYGEYRVVVVSVG 5779
Qy 427 -----LDRI----- 430
Db 5780 SDSGTVNTVADVDPPEPPPIIENILDEAVILSWKPPALDGGSLVTNVTIEKREAMGGSW 5839
Qy 431 ----- 430
Db 5840 SPCAKSRYTYTTIEGLRAGQYEFRIIAENKHGSKPCCEPTAPVLIPGDERKRRRGYDVD 5899
Qy 431 ----- 430
Db 5900 EGGKIVRGKGTSSNYDNVFDIWKQYYPQPVBIKHDHVLDRHYDIHEELGTGAFGVVHRV 5959
Qy 431 -----RNFS-----LRSPYYNL----- 443
Db 5960 TERATGNNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAPEDDNEWMYIYEFM 6019
Qy 444 ----- 443
Db 6020 SGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNYVHLDLKPENIMFTTKRSNEL 6079
Qy 444 ----- 443
Db 6080 KLIDPGLTAHLDPKQSVKVTGTAEFAAPEVAEGKPGVGYTDMMSVGVLSYLLSGLSPF 6139
Qy 444 ----- 443
Db 6140 GGENDETIRNVKSCDWMDDSAFSGISEDGKDFIRKLLADPNTRMTIHOALEHPWLTP 6199
Qy 444 -----Tpp-----EHR----- 449
Db 6200 GNAPGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGRISNYSSLRKHQRPQEYSIRDAPW 6259
Qy 450 ----- 449
Db 6260 DRSEAQRPIVVKPYGTEVGEQSANFYCRVYIASSPPVTVHKKDRELKQSVKMKRYNGN 6319
Qy 450 -----RAVN----- 453
Db 6320 DYGLTINRVKDDKGEYTVRAKNISYGTKEIVFLNVTRHSBPLKPEPLEPMKAPSPRV 6379
Qy 454 ----- 453
Db 6380 EEFKERSAPPTFHLNRLLIQNHCKLCTCSLOGNPNTIEWMKDGHVPDEDRVQVSFR 6439
Qy 454 ----- 453
Db 6440 SGVCSLEIFNARVDDAGTYTATNDLGVDSVCVLTVQTKGGEPIPRVSSFRPRRAYDT 6499
Qy 454 LST----- 456
Db 6500 LSTGTDVERSHYADMRRLIRDVSPDVSAAADDLTKITNELPSPFTAQLSDSETEVGG 6559
Qy 457 ----- 456
Db 6560 SAEFSAAVSGQPEPLIEWLHNGERISEDSRFRASYVAGKATLRISDAKXDEGOYL CRA 6619
Qy 457 SNS----- 459
Db 6620 SNSAQEQTRATLVKGDQPLNGHAGQAVSELVRVKHLGGEIVNNGESVTFEARVQGT 6679
Qy 460 -----LW----- 461
Db 6680 PEEVLWNRNGOELTNGDKTISQDGETLSFTINSADASDAGHYQLEVRKGTNLVSVASL 6739
Qy 462 -----WV 463
Db 6740 VVVGKADPPVTRLPSSVSAPLGGSTAFTIEFENVEGLTVQWP 6782

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:22:26 ; Search time 196 Seconds
(without alignments)
1365.047 Million cell updates/sec

Title: US-09-942-146A-1
Perfect score: 2451
Sequence: 1 MGRKEMVVRDPKMFVLISI.....PEHRRVNLSTSNLWMLQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.0

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2417	98.6	466	1 UL74 HCMVA	P16750 human cytom
2	2417	98.6	466	2 Q7M618	Q7M618 human cytom
3	2417	98.6	466	2 AAN40055	AAN40055 human cyto
4	2411	98.4	466	2 Q8AZ28	Q8AZ28 human cytom
5	2407	98.2	466	2 Q8BCU8	Q8BCU8 human cytom
6	2398	97.8	466	2 Q8AZ36	Q8AZ36 human cytom
7	2398	97.8	466	2 Q8BCU2	Q8BCU2 human cytom
8	2071	84.5	463	2 Q8AZ32	Q8AZ32 human cytom
9	2069	84.4	464	2 Q8BCU3	Q8BCU3 human cytom
10	2068	84.4	463	2 Q8BCU9	Q8BCU9 human cytom
11	2068	84.4	463	2 Q8BCV2	Q8BCV2 human cytom
12	2067	84.3	463	2 Q8AZ39	Q8AZ39 human cytom
13	2057	83.9	464	2 Q8BCU8	Q8BCU8 human cytom
14	2057	83.9	464	2 AAS48965	AAS48965 human cyto
15	2054	83.8	462	2 Q8BCU0	Q8BCU0 human cytom
16	2049	83.6	462	2 Q8BCU6	Q8BCU6 human cytom
17	1947	79.4	465	2 Q8AYW0	Q8AYW0 human cytom
18	1941	79.2	465	2 Q8BCU9	Q8BCU9 human cytom
19	1906	77.8	462	2 Q8AYZ3	Q8AYZ3 human cytom
20	1901	77.6	462	2 Q8BCU5	Q8BCU5 human cytom
21	1877	76.6	472	2 Q8BCV1	Q8BCV1 human cytom
22	1877	76.6	472	2 AAR31626	AAR31626 human cyto
23	1759	71.8	464	2 Q8BCU1	Q8BCU1 human cytom
24	1751	71.4	464	2 Q8BCU4	Q8BCU4 human cytom
25	1748	71.3	464	2 Q8AYV5	Q8AYV5 human cytom
26	1746	71.2	464	2 Q8AYZ5	Q8AYZ5 human cytom
27	1740	71.0	457	2 Q8BCU7	Q8BCU7 human cytom
28	1740	71.0	457	2 Q8BCV0	Q8BCV0 human cytom
29	1026	41.9	461	2 Q8QS25	Q8QS25 chimpanzee
30	820	33.5	389	2 Q7TFN1	Q7TFN1 thesus cyto
31	718	29.3	144	2 Q7T9B3	Q7T9B3 human cytom

ALIGNMENTS

RESULT 1
UL74 HCMVA

ID UL74 HCMVA STANDARD; PRT; 466 AA.

AC P16750;

DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Glycoprotein UL74 precursor.

GN Name=UL74;

OS Human cytomegalovirus (strain AD169).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI_TaxID=10360;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=90269039; PubMed=2161319;

RA Chee N.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,

RA Horsnell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,

RA Preddie B., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;

RT "Analysis of the protein-coding content of the sequence of human

cytomegalovirus strain AD169."

RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC -1- SIMILARITY: Belongs to the herpesviruses U47 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X17403; CAA35389.1; -.

DR PIR; S09837; S09837.

KW Glycoprotein; Signal.

FT SIGNAL 1 30

FT CHAIN 31 466 Potential.

FT CARBOHYD 75 75 Glycoprotein UL74.

FT CARBOHYD 83 83 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 87 87 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 103 103 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 130 130 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 157 157 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 162 162 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 171 171 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 219 219 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 242 242 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 288 288 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 292 292 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 350 350 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 385 385 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 392 392 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 399 399 N-linked (GlcNAc. .) (Potential).

FT CARBOHYD 433 433 N-linked (GlcNAc. .) (Potential).

32 718 29.3 144 2 Q7T9B4 human cytom

33 718 29.3 144 2 Q7T9B5 human cytom

34 718 29.3 144 2 Q7T9B6 human cytom

35 718 29.3 144 2 Q7T9C3 human cytom

36 718 29.3 144 2 Q7T9C4 human cytom

37 718 29.3 144 2 Q7T9C5 human cytom

38 718 29.3 144 2 Q7T9D2 human cytom

39 718 29.3 144 2 Q7T9D4 human cytom

40 718 29.3 144 2 Q7T9D5 human cytom

41 711 29.0 34350 2 Q8WZ42 homo sapien

42 707 28.8 144 2 Q7T9B2 human cytom

43 683 27.9 144 2 Q7T9C1 human cytom

44 681 27.8 26926 2 Q10466 homo sapien

45 680 27.7 20925 2 Q8N0X1 trichoderma

FT CARBOHYD 454 454 N-linked (GlcNAc....) (Potential).
 SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 98.6%; Score 2417; DB 1; Length 466;
 Best Local Similarity 98.9%; Pred. No. 8.8e-58;
 Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
 DB 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60

QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
 DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120

QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179
 DB 121 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 180

QY 180 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVMSRNLFRVPKY 239
 DB 181 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVMSRNLFRVPKY 240

QY 240 INGTKLNTWRKLRKQAPVKEQEKAKTKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 299
 DB 241 INGTKLNTWRKLRKQAPVKEQEKAKTKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 300

QY 300 ARRVSTSTIAYRPDSSFMKSIATOLRDLATWYTTLYRQNPFCPEPSRNTAVSEFMKN 359
 DB 301 ARRVSTSTIAYRPDSSFMKSIATOLRDLATWYTTLYRQNPFCPEPSRNTAVSEFMKN 360

QY 360 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
 DB 361 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420

QY 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWWLQ 465
 DB 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWWLQ 466

RESULT 2
 Q7M6L8 PRELIMINARY; PRT; 466 AA.

ID Q7M6L8
 AC Q7M6L8;
 DT 01-MAR-2004 (TremBLrel. 26, Created)
 DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
 DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
 DE UL74.
 OS Human cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ad169;
 RX MEDLINE=22421467; PubMed=12533697;
 RA Davidson A.J., Dolan A., Akter P., Addison C., Dargan D.J.,
 RA Alexander D.J., McGeoch D.J., Hayward G.S.;
 RT "The human cytomegalovirus genome revisited: comparison with the
 RT chimpanzee cytomegalovirus genome.";
 RL J. Gen. Virol. 84:17-28(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ad169;
 RX MEDLINE=22255414; PubMed=12368327;
 RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
 RA "The genes encoding the gCIII complex of human cytomegalovirus exist
 RT in highly diverse combinations in clinical isolates.";
 RL J. Virol. 76:10841-10848(2002).
 DR EMBL; BK000394; DAA00170.1; -.
 DR EMBL; AF531331; AAN40055.1; -.
 SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 98.6%; Score 2417; DB 2; Length 466;
 Best Local Similarity 98.9%; Pred. No. 8.8e-58;
 Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
 DB 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60

QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
 DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120

QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179
 DB 121 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 180

QY 180 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVMSRNLFRVPKY 239
 DB 181 NVPRWNTKLYVGPVKVNDVSDTIYFLGLTALLRYAQRNCTHSFYLNVMSRNLFRVPKY 240

QY 240 INGTKLNTWRKLRKQAPVKEQEKAKTKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 299
 DB 241 INGTKLNTWRKLRKQAPVKEQEKAKTKTOSTTTPYPSYTTSAALNVTTNVTYSITTA 300

QY 300 ARRVSTSTIAYRPDSSFMKSIATOLRDLATWYTTLYRQNPFCPEPSRNTAVSEFMKN 359
 DB 301 ARRVSTSTIAYRPDSSFMKSIATOLRDLATWYTTLYRQNPFCPEPSRNTAVSEFMKN 360

QY 360 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 419
 DB 361 THVLIRNETPYTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFQRTFIDPLWD 420

QY 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWWLQ 465
 DB 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWWLQ 466

RESULT 3
 AAN40055 PRELIMINARY; PRT; 466 AA.

ID AAN40055
 AC AAN40055;
 DT 02-MAR-2004 (TremBLrel. 27, Created)
 DT 02-MAR-2004 (TremBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TremBLrel. 27, Last annotation update)
 DE UL74 protein.
 OS Human cytomegalovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ad169;
 RX MEDLINE=22255414; PubMed=12368327;
 RA Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;
 RA "The Genes Encoding the gCIII Complex of Human Cytomegalovirus Exist
 RT in Highly Diverse Combinations in Clinical Isolates.";
 RL J. Virol. 76:10841-10848(2002).
 DR EMBL; AF531331; AAN40055.1; -.
 SQ SEQUENCE 466 AA; 54234 MW; 20B931D97AB5D41D CRC64;

Query Match 98.6%; Score 2417; DB 2; Length 466;
 Best Local Similarity 98.9%; Pred. No. 8.8e-58;
 Matches 461; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
 DB 1 MGRKEMVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60

QY 61 EILRQLETTISTKYNVSKQPVKNLTWN-TEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 119
 DB 61 EILRQLETTISTKYNVSKQPVKNLTWNTEFPQYIILAGPIQNYISITYLWFDYFYSTQLRK 120

QY 120 PAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEOGCGNFTTFNPMFF 179


```

Db 121 PAKVYVSYNHAKTITFRPPCGTVPMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFORFTIDPLWD 419
Db 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

RESULT 4

```

Q8AZ28 ID Q8AZ28 PRELIMINARY; PRT; 466 AA.
AC Q8AZ28;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=298, 6444, and SW3;
RX MEDLINE=22255414; PubMed=12368327;
RA Raemussen L., Geisler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
RL in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531320; AAN40044.1; -
DR EMBL; AF531325; AAN40049.1; -
DR EMBL; AF531337; AAN40061.1; -
DR EMBL; AF531346; AAN40070.1; -
SQ SEQUENCE 466 AA; 54200 MW; 56581B75919572FE CRC64;

```

```

Query Match 98.4%; Score 2411; DB 2; Length 466;
Best Local Similarity 98.7%; Pred. No. 1.3e-57;
Matches 460; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 1 MGRKEMVVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEMVVRDVPKMFVLISISFLVSLINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTMN-TEPPQYIILAGPIQNYISITVLPDFYSTOLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTMNTEPPQYIILAGPIQNYISITVLPDFYSTOLRK 120
Qy 120 PAKVYVSYNHAKTITFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSYNHAKTITFRPPCGTVPMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 EILRQLETTISTKYNVSKQPVKNLTMNTEPPQYIILAGPIQNYISITVLPDFYSTOLRK 120
Qy 120 PAKVYVSYNHAKTITFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSYNHAKTITFRPPCGTVPMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300

```

```

Qy 300 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFORFTIDPLWD 419
Db 361 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFORFTIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

RESULT 5

```

Q8BCU8 ID Q8BCU8 PRELIMINARY; PRT; 466 AA.
AC Q8BCU8;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=851;
RX MEDLINE=22255414; PubMed=12368327;
RA Raemussen L., Geisler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
RL in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531330; AAN40054.1; -
SQ SEQUENCE 466 AA; 54230 MW; 56581B75919562FF CRC64;

```

```

Query Match 98.2%; Score 2407; DB 2; Length 466;
Best Local Similarity 98.5%; Pred. No. 1.7e-57;
Matches 459; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 MGRKEMVVRDVPKMFVLISISFLVSPINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEMVVRDVPKMFVLISISFLVSLINCKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Qy 61 EILRQLETTISTKYNVSKQPVKNLTMN-TEPPQYIILAGPIQNYISITVLPDFYSTOLRK 119
Db 61 EILRQLETTISTKYNVSKQPVKNLTMNTEPPQYIILAGPIQNYISITVLPDFYSTOLRK 120
Qy 120 PAKVYVSYNHAKTITFRPPCGRVPSTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 179
Db 121 PAKVYVSYNHAKTITFRPPCGTVPMTCLSEMLNVSKNDTGEQCGNFTTFNPMFF 180
Qy 180 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 239
Db 181 NVPRWNTKLYVGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNAMSRLNFRVPKY 240
Qy 240 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
Db 241 INGTKLKNTWKLKRRKQAPVKEQEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
Qy 300 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 359
Db 301 ARRVSTSTIAYRPDSSPFMKSIATQRLDRLATWVYTLRYRONPFCPSRNRATAVSEFMKN 360
Qy 360 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFORFTIDPLWD 419
Db 361 THVLIRNETPTIYGTLDMSLYNETMFVENKTASDSNKTTPSPSGMGFORFTIDPLWD 420
Qy 420 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 465
Db 421 YLDSLLFLDEIRNFSRLSPYVNLTPPEHRRAVNLSTNSLWMLQ 466

```

```
RESULT 6
Q8AZ36
ID Q8AZ36 PRELIMINARY; PRT; 466 AA.
AC Q8AZ36;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC9558443E365A0 CRC64;

Query Match 97.8%; Score 2398; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 3e-57;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
DB 1 MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
QY 61 EILRQLETTISTKYNVSKQVKNLTWN-TEFPQYIILAGPIQNYSTIYLWDFYSTQLRK 119
DB 61 EILRQLETTISTKYNVSKQVKNLTWNTEFPQYIILAGPIQNYSTIYLWDFYSTQLRK 120
QY 120 PAKYVVSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVVSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 239
DB 181 NVPRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 240
QY 240 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
DB 241 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
QY 300 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRATAVSEFMKN 359
DB 301 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRATAVSEFMKN 360
QY 360 THVLIRNETPPTYIYGTLDMSLSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLWD 419
DB 361 THVLIRNETPPTYIYGTLDMSLSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 7
Q8BCU2
ID Q8BCU2 PRELIMINARY; PRT; 466 AA.
AC Q8BCU2;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4088, and SW1762;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531322; AAN40046.1; -.
DR EMBL; AF531344; AAN40068.1; -.
SQ SEQUENCE 466 AA; 54215 MW; 4DC9558443E365A0 CRC64;

Query Match 97.8%; Score 2398; DB 2; Length 466;
Best Local Similarity 98.3%; Pred. No. 3e-57;
Matches 458; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
DB 1 MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRGLVLSKIGKYLQDLKL 60
QY 61 EILRQLETTISTKYNVSKQVKNLTWN-TEFPQYIILAGPIQNYSTIYLWDFYSTQLRK 119
DB 61 QILRQLETTISTKYNVSKQVKNLTWNTEFPQYIILAGPIQNYSTIYLWDFYSTQLRK 120
QY 120 PAKYVVSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 179
DB 121 PAKYVVSQYNHTAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNPMFF 180
QY 180 NVPRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 239
DB 181 NVPRWNTKLYVGPTKVNVDSTQIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPKY 240
QY 240 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 299
DB 241 INGTKLNTMRKLKRQAPVKEQFEKAKKTQSTTTPYFSYTTSAALNVTNTVYSITTA 300
QY 300 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRATAVSEFMKN 359
DB 301 ARRVSTTIAYRPDSFMSKINATQRLDRLATWVYTLRYRONPCEPSRNRATAVSEFMKN 360
QY 360 THVLIRNETPPTYIYGTLDMSLSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLWD 419
DB 361 THVLIRNETPPTYIYGTLDMSLSLYNETMFVENKTASDNKTTPTSPSMGFORTFIDPLWD 420
QY 420 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 465
DB 421 YLDSLLFLDEIRNFSLSRSPTYVNLTPPEHRAVNLSTNSLWMLQ 466

RESULT 8
Q8AZ32
ID Q8AZ32 PRELIMINARY; PRT; 463 AA.
AC Q8AZ32;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=279, and SW2;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates.";
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531319; AAN40043.1; -.
DR EMBL; AF531345; AAN40069.1; -.
SQ SEQUENCE 463 AA; 53871 MW; 51310C0D1D239A5A CRC64;
```

Query Match	84.5%;	Score 2071;	DB 2;	Length 463;	
Best Local Similarity	85.4%;	Pred. No. 3.4e-48;			
Matches 399;	Conservative 22;	Mismatches 40;	Indels 6;	Gaps 4	

Qy	1	MGRKEMVMDVPKMFVLISISFLVSPINCKVMSKALYNRPWRLGLVLSKIGKYKLDQLKL	60
Db	1	MGRKEM- RGVNLFPLMSLTFLLPSFINCRAAVRLSVGRVWSGKVLSTIGQRUDKEFL	59
Qy	61	EILRLQLETTISTKY - NVSKQVPKQLTWN - TEPPQYIILAGPIQNYSIYILWFDYPYSTQLR	118
Db	60	EILKQLEKDIYTKYFNMTQHIKNLTWMTSEPPYIILAGPIQNNSVTYLWFDYPYSTQLR	119
Qy	119	KPAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTFNMP	178
Db	120	KPAKYVSEYNHTAKTITFRPPSCGTVPSTCLSEMLNVSKRNDTGEQCGNFTTFNMP	179
Qy	179	FNVPRWNTKLVGPTKVNVDSTIYFGLGTALLRYAQRNCTHSPYLNVMSRNLFRVPK	238
Db	180	FNVPRWNKLVGSKKVNVDSTIYFGLGTALLRYAQRNCTHSPYLNVMSRNLFRVPK	239
Qy	239	YINGTKLKNMRKLKRQAQPVKEQFEKKAQKTSQTTTPYFYSYTSAAALNVTTNVTYSITT	298
Db	240	YINGTKLKNMRKLKRQAQPVKEQSEKSKSQSTTPYPSPYTTSTALNVTTNATYSVTT	299
Qy	299	AARVSTSTIAYRPDSSPFMSKIMATQLADLATWYTLRYQNPFCEPSRNRATVSEPMK	358
Db	300	TARVSTSTIAYRPDSSPFMSKIMTQLADLATWYTLRYQNPFCESSRNRATVSEPMK	359
Qy	359	NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW	418
Db	360	NTHVLIRNETPYTYIGTLDSSLYNETMFVENKTASE---TTPSPSTGFQRTFIDPLW	416
Qy	419	DYLDLSLLFLDEIRNFSRSPTYNLTTPPEHRRVNLSTNSLWMLQ	465
Db	417	DYLDLSLLFLDEIRNFSLSQSPYGNLTTPPEHRRVNLSTNSLWMLQ	463

RESULT 9					
Q8BCU3	Q8BCU3	PRELIMINARY;	PTT;	464 AA.	
AC	Q8BCU3				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				
DE	UL74 protein.				
OS	Human cytomegalovirus.				
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;				
OC	Betaherpesvirinae; Cytomegalovirus.				
OX	NCBI_TaxID=10359;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SW1324;				
RC	MEDLINE=22255414; PubMed=12368327;				
RA	Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;				
RT	"The genes encoding the gCIII complex of human cytomegalovirus exist				
RT	in highly diverse combinations in clinical isolates.";				
RL	J. Virol. 76:10841-10848(2002).				
DR	EMBL: AF531340; AAN40064.1;				
SQ	SEQUENCE 464 AA; 54146 MW; A24A513BF112848B CRC64;				

Query Match	84.4%;	Score 2069;	DB 2;	Length 464;	
Best Local Similarity	84.3%;	Pred. No. 3.9e-48;			
Matches 397;	Conservative 25;	Mismatches 36;	Indels 13;	Gaps 5;	

Qy	1	MGRKEMVMDVPKMFVLISISFLVSPINCKVMSKALYNRP-----WRGLVLSTKIGKYKLD	56
Db	1	MGRKEDN-RSISKLFFIISLTVLFSINCKV-----RPPGRYWLGTVLSTIGKQKLD	53
Qy	57	QLKLEILRLQLETTISTKY - NVSKQVPKQLTWN - TEPPQYIILAGPIQNYSIYILWFDYPS	114
Db	54	KFKLEILKQLEREPYTYFNTRQVRKRLTNWMTQFPQYIILAGPIRNDSTIYLWFDYPS	113
Qy	115	TQLRKPAKYVYSQYNHTAKTITFRPPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTTF	174

Db	114	TQURKPAKYYSQYNHTAKTITFRPPSCGTVPSMTCISEMLNVSKRNDTGSGCGNFTTF	173
Qy	175	NPMPFNVRNWKLYVGPTKVNVDSTQIYFLGLTALLRLYAQRNCTHSFYLVNMSRNL	234
Db	174	NPMPFNVRNWKLYVGPTKVNVDSTQIYFLGLTALLRLYAQRNCTHSFYLVNMSRNL	233
Qy	235	RVPKYINGTKLKNTRMKLKRQAPVKQSEKKAKTQSTTTPYFSYTTSAALNVTNVTY	294
Db	234	RVPKYINGTKLKNTRMKLKRQAPVKQSEKKTQSTTTPYFSYTTSTALNVTNATY	293
Qy	295	SIITAAARVSTSIAYRPDSSFMSKINATQRLDIAWYTTTLRYQRNPFCEPSNRRTAVS	354
Db	294	RVTTSAKRIPSTIAYRPDSSFMSKINATQRLDIAWYTTTLRYRNEPFCCKPDRNRRTAVS	353
Qy	355	EFKNTHTVLNRNETPYIYGTLDMSLYINETMVENKTASDSNKTTPTSPSMGFORTFI	414
Db	354	EFKNTHTVLNRNETPYIYGTLDMSLYINETHSVENETASDNNETPTSPSTFQKTFI	413
Qy	415	DPLWDYDLSLLFLDEIRNPSLRGTYVNLTPPEHRRANVLSTNSLMMWLQ	465
Db	414	DPLWDYDLSLLFLDKIRNPSLQPAYGNLTPEHRRANVLSTNSLMMWLQ	464
RESULT 10			
Qy	Q8BCT9	PRELIMINARY; PRT; 463 AA.	
AC	Q8BCT9;		
DT	01-MAR-2003 (T-EMBLrel. 23, Created)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last sequence update)		
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)		
DE	UL74 protein.		
OS	Human cytomegalovirus.		
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae;		
OC	Betaherpesvirinae; Cytomegalovirus.		
OX	NCBI_TaxID=10359;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SW990;		
RX	MEDLINE=22255414; PubMed=12368327;		
RA	Rasmussen L., Geisler A., Cowan C., Chase A., Winters M.;		
RT	"The genes encoding the gCIII complex of human cytomegalovirus exist		
RL	in highly diverse combinations in clinical isolates."		
RL	J. Virol. 76:10841-10848(2002).		
DR	EMBL; AF531354; AAM40078.1; -.		
SQ	SEQUENCE 463 AA; 53885 MW; B86783A05FB9424A CRC64;		
Query Match 84.4%; Score 2068; DB 2; Length 463;			
Best Local Similarity 85.2%; Pred. No. 4.1e-48;			
Matches 398; Conservative 23; Mismatches 40; Indels 6; Gaps 4;			
Qy	1	MGRKEMVRDVPKMFVLISIFLLVSPINCKVMSKALYNRPWRLGLVLSKIGKYLQDLKL	60
Db	1	MGRKGEV-RGVFLFFLMTITFLFSPINCRVAVRLSVGRVWSGVJSTIGKQLDKPKL	59
Qy	61	EILRLQLETTISTKY-NVSKOPVKNLTNW-TEFPQYIILAGPIQNYSTIYLMFDFYSTQLR	118
Db	60	EILKQLEKDIYTKYFNMTROHINKLNTWMTFFRYIILAGPIQNNSTYLMFDFYSTQLR	119
Qy	119	KPAKYVYSQYNHTAKTITFRPPSCGTVPSMTCISEMLNVSKRNDTGSGCGNFTTFNPMF	178
Db	120	KPAKYVSEYNHTAKTITFRPPSCGTVPSMTCISEMLNVSKRNDTGSGCGNFTTFNPMF	179
Qy	179	FNVPRNWKLYVGPTKVNVDSTQIYFLGLTALLRLYAQRNCTHSFYLVNMSRNLFRVPK	238
Db	180	FNVPRNWKLYVGSKKVNVDSTQIYFLGLTALLRLYAQRNCTHSFYLVNMSRNLFRVPK	239
Qy	239	YINGTKLKNTRMKLKRQAPVKQSEKKAKTQSTTTPYFSYTTSAALNVTNVTY	298
Db	240	YINGTKLKNTRMKLKRQAPVKQSEKSKKQSTTTPYSPYTTSTALNVTNATY	299
Qy	299	AARVSTSTIAYRPDSSFMSKINATQRLDIAWYTTTLRYQRNPFCEPSNRRTAVSEFMK	358

```

Db 300 TARRVSTSTIAYRDPSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASE--TTPTSPSTGFGRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

RESULT 11
Q8BCV2
ID Q8BCV2 PRELIMINARY; PRT; 463 AA.
AC Q8BCV2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=122;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates."
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531316; AAN40040.1; -.
DR EMBL; AF531316; AAN40040.1; -.
SQ SEQUENCE 463 AA; 53957 MW; 55A7D4157FD32424 CRC64;
[1]
Query Match 84.4%; Score 2068; DB 2; Length 463;
Best Local Similarity 85.4%; Pred. No. 4.1e-48;
Matches 399; Conservative 22; Mismatches 40; Indels 6; Gaps 4;
Qy 1 MGRKEMVRDVPKMFVLISISLVLVGFINKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEM-RGVNLFLLMSLTLFLFSPINCAARVLSVGRVMSGKVLSTIGKRLDKPKL 59
Qy 61 EILRQLETTSTKY-NVSKOPVKNLTWN-TEFPQYVILAGPIQNYSTVLMFDFYSTOLR 118
Db 60 EILKQLEKDIYTKYFNMTQHINKLTMNTEFPRIYVILAGPIQNNSTVLMFDFYSTOLR 119
Qy 119 KPAKYVSOVNHAKTITRPPCGRVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEVNHAKTITRPPCGRVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 179
Qy 179 FNVPRNNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRNNTKLYGSKKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239
Qy 239 YINGTKLNTMKLRKQAPVKEQEKAKTQSTTTPYFYSYTTSAALNVTNTVYSITT 298
Db 240 YINGTKLNTMKLRKQAPVKEQEKSKQSQTTPYSPYTTSTALNVTNTVYSITT 299
Qy 299 AARRVSTSTIAYRDPSSFMKSIWTTQLRDLATWYVTLRYRQNPFCPSNRRTAVSEFMK 358
Db 300 TARRVSTSTIAYRDPSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASE--TTPTSPSTGFGRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

```

RESULT 12
Q8AZ39

```

ID Q8AZ39 PRELIMINARY; PRT; 463 AA.
AC Q8AZ39;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=650, and DM7;
RX MEDLINE=22255414; PubMed=12368327;
RA Rasmussen L., Geissler A., Cowan C., Chase A., Winters M.;
RT "The genes encoding the gCIII complex of human cytomegalovirus exist
in highly diverse combinations in clinical isolates."
RL J. Virol. 76:10841-10848(2002).
DR EMBL; AF531326; AAN40050.1; -.
DR EMBL; AF531334; AAN40058.1; -.
SQ SEQUENCE 463 AA; 53901 MW; 403100AADBF4412A CRC64;
[1]
Query Match 84.3%; Score 2067; DB 2; Length 463;
Best Local Similarity 85.2%; Pred. No. 4.4e-48;
Matches 398; Conservative 22; Mismatches 41; Indels 6; Gaps 4;
Qy 1 MGRKEMVRDVPKMFVLISISLVLVGFINKVMSKALYNRPWRGLVLSKIGYKLDQLKL 60
Db 1 MGRKEM-RGVNLFLLMSLTLFLFSPINCAARVLSVGRVMSGKVLSTIGKRLDKPKL 59
Qy 61 EILRQLETTSTKY-NVSKOPVKNLTWN-TEFPQYVILAGPIQNYSTVLMFDFYSTOLR 118
Db 60 EILKQLEKDIYTKYFNMTQHINKLTMNTEFPRIYVILAGPIQNNSTVLMFDFYSTOLR 119
Qy 119 KPAKYVSOVNHAKTITRPPCGRVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 178
Db 120 KPAKYVSEVNHAKTITRPPCGRVPSMTCLSEMLNVSKNDTGEQCGNFTTFNPMF 179
Qy 179 FNVPRNNTKLYGPTKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 238
Db 180 FNVPRNNTKLYGSKKVNVDSTIYFLGLTALLRYAQRNCTHSPYLVNMSRNLFRVPK 239
Qy 239 YINGTKLNTMKLRKQAPVKEQEKAKTQSTTTPYFYSYTTSAALNVTNTVYSITT 298
Db 240 YINGTKLNTMKLRKQAPVKEQEKSKQSQTTPYSPYTTSTALNVTNTVYSITT 299
Qy 299 AARRVSTSTIAYRDPSSFMKSIWTTQLRDLATWYVTLRYRQNPFCPSNRRTAVSEFMK 358
Db 300 TARRVSTSTIAYRDPSSFMKSIWTTQLRDLATWYVTLRYRQNPFCSSRNRTAVSEFMK 359
Qy 359 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASDNKTTPTSPSMGFQRTFIDPLW 418
Db 360 NTHVLIRNETPYTYIGTLDMSLLYNETMFVENKTASE--TTPTSPSTGFGRTFIDPLW 416
Qy 419 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 465
Db 417 DYLDLFLDLIRNFSRLSPYVNLTPPEHRAVNLTSTNSLWMLQ 463

RESULT 13
Q8BCT8
ID Q8BCT8 PRELIMINARY; PRT; 464 AA.
AC Q8BCT8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE UL74 protein.
GN Name=UL74; ORFName=HHV5g069;
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]

```


Db 60 EILKQLEKDIYKYFNMTQHIKNLTWNMTPEPRYYILAGPIQNNSVTYLVWDFYSTQLR 119
Qy 119 KPAKYVYSQVNHAKTITFRPPCGRVPSMTCLSEMLNVSKRNDTGEQCGNFTFNPMF 178
Db 120 KPAKYVSEYNHAKTITFRPPCGTVPSMTCLSEMLNVSKRNDTGEQCGN-TTFNPMF 178
Qy 179 FNVRWNTKLYVGTQKVNDSQTIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPK 238
Db 179 FNVRWNTKLYVGSKKVNDSQTIYFLGLTALLRYAQRNCTHSFVLNAMSRLFRVPK 238
Qy 239 YINGTKLKNTRKLRKQAPVKEOPEKAKKQSTTTPYFSYTTSAALNVTTNVTYSITT 298
Db 239 YINGTKLKNTRKLRKQAPVKEQSEKSKSQSTTTPYSPYTTSTALNVTTNATYSVTT 298
Qy 299 AARRVSTSTIAYRPDSSFMSKIMATQRLDRLATWVTTILRYRONPCEPSRNRNAVSEFMK 359
Db 299 TARRISTSTIAYRPDSSFMSKIMTTQLRLDRLATWVTTILRYRONPCESSRNRNAVSEFMK 358
Qy 359 NTHVLIRNETPYTIYGTLDMSLYYNETMFVENKTASDNKTTTPTSPSMGFQRTFIDPLW 418
Db 359 NTHVLIRNETPYTIYGTLDMSLYYNETMFVENKTASE---TTPTSPSTGFQRTFIDPLW 415
Qy 419 DYLDLSLLFLDEIRNFSLRPTVNLTPPEHRAVNLSNSTSLWWWLQ 465
Db 416 DYLDLSLLFLDEIRNFSLOSPTVGNLTPEHRAVNLSNSTSLWWWLQ 462

Search completed: December 7, 2004, 15:39:25
Job time : 200 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 7, 2004, 15:30:25 ; Search time 40 Seconds
(without alignments)
770.947 Million cell updates/sec

Title: US-09-942-146A-1
Perfect score: 2451
Sequence: 1 MGRKEMVDRVPMFVLISI.....PEHRRVNLSTNSLWMLQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.0

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/prodata/1/iaa/5A COMB pep.*
2: /cgn2_6/prodata/1/iaa/5B COMB pep.*
3: /cgn2_6/prodata/1/iaa/6A COMB pep.*
4: /cgn2_6/prodata/1/iaa/6B COMB pep.*
5: /cgn2_6/prodata/1/iaa/PCUTUS COMB pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2420	98.7	465	US-09-627-986-1	Sequence 1, Appli
2	575	23.5	15281	US-08-471-119A-2	Sequence 2, Appli
3	516	21.1	10182	US-09-134-001C-3159	Sequence 3159, Ap
4	483	19.7	11877	US-09-105-537-6	Sequence 6, Appli
5	449	18.3	4563	US-09-108-006C-1	Sequence 1, Appli
6	449	18.3	4563	US-09-538-092-842	Sequence 842, App
7	440	18.0	4536	US-09-180-422B-27	Sequence 27, Appli
8	440	18.0	4536	US-09-079-030-1	Sequence 1, Appli
9	426	17.4	3898	US-08-750-717-2	Sequence 2, Appli
10	422	17.2	3778	US-08-222-617A-2	Sequence 2, Appli
11	420	17.1	3666	US-08-222-617A-12	Sequence 12, Appli
12	420	17.1	3727	US-08-222-617A-27	Sequence 27, Appli
13	420	17.1	3898	US-08-876-991-2	Sequence 2, Appli
14	420	17.1	3898	US-09-059-853-2	Sequence 2, Appli
15	417	17.0	4654	US-08-476-515A-84	Sequence 84, Appli
16	416	17.0	4655	US-08-652-877-84	Sequence 84, Appli
17	416	17.0	4655	US-08-652-877-86	Sequence 86, Appli
18	416	17.0	4655	US-08-652-877-88	Sequence 88, Appli
19	416	17.0	4655	US-08-652-877-90	Sequence 90, Appli
20	412	16.8	3623	US-09-341-461-2	Sequence 2, Appli
21	411	16.8	7257	US-09-335-409-5	Sequence 5, Appli
22	411	16.8	7257	US-09-568-102-5	Sequence 5, Appli
23	411	16.8	7257	US-09-567-969-5	Sequence 5, Appli
24	411	16.8	7257	US-09-568-480-5	Sequence 5, Appli
25	411	16.8	7257	US-09-568-486-5	Sequence 5, Appli
26	411	16.8	7257	US-09-568-472-5	Sequence 5, Appli
27	411	16.8	7257	US-09-567-899-5	Sequence 5, Appli

28	409	16.7	3665	2	US-08-222-617A-13	Sequence 13, Appli
29	409	16.7	3712	2	US-08-222-617A-4	Sequence 4, Appli
30	409	16.7	3712	2	US-08-222-617A-25	Sequence 25, Appli
31	409	16.7	5032	4	US-09-538-092-979	Sequence 2, Appli
32	408	16.6	3461	3	US-09-334-220-2	Sequence 2, Appli
33	406	16.6	3969	3	US-08-061-376-5	Sequence 5, Appli
34	405	16.5	3869	4	US-09-538-092-1262	Sequence 1262, Ap
35	405	16.5	4861	4	US-09-919-497-70	Sequence 70, Appli
36	403	16.4	4866	3	US-09-424-783-2	Sequence 2, Appli
37	402	16.4	3460	3	US-09-334-220-1	Sequence 1, Appli
38	401	16.4	5405	3	US-08-718-388-9	Sequence 9, Appli
39	396	16.2	3959	2	US-08-970-269A-30	Sequence 30, Appli
40	386	16.2	3959	3	US-09-407-562-30	Sequence 30, Appli
41	394	16.1	5588	3	US-09-036-987A-6	Sequence 6, Appli
42	394	16.1	5588	3	US-09-370-700-6	Sequence 6, Appli
43	389	15.9	3443	2	US-09-603-207-6	Sequence 6, Appli
44	389	15.9	3443	2	US-08-416-603-2	Sequence 2, Appli
45	389	15.9	6396	4	US-09-410-551B-72	Sequence 72, Appli

ALIGNMENTS

RESULT 1
US-09-627-986-1
; Sequence 1, Application US/09627986
; Patent No. 6569616
; GENERAL INFORMATION:
; APPLICANT: Compton, Teresa
; APPLICANT: Huber, Mary T.
; TITLE OF INVENTION: HUMAN CYTOMEGALOVIRUS GLYCOPROTEIN O AS A NEW DRUG
; FILE REFERENCE: 960296.96579
; CURRENT APPLICATION NUMBER: US/09/627,986
; CURRENT FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: 60/146,180
; PRIOR FILING DATE: 1999-07-29
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Human cytomegalovirus
US-09-627-986-1

Query Match		98.7%	Score 2420;	DB 4;	Length 465;
Best Local Similarity		99.4%	Pred. No. 4.3e-72;		
Matches 463;		Conservative 0;	Mismatches 1;	Indels 2;	Gaps 2;
Qy	1	MGRKEMVDRVPMFVLISISFLLVSVFINKVMSKALYNRPWGLVLSKIGKYKLDQLKL	60		
Db	1	MGRKEMVDRVPMFVLISISFLLVSVFINKVMSKALYNRPWGLVLSKIGKYKLDQLKL	60		
Qy	61	EILRQLETTISTIKYNSKOPVKNLTNTWTFPPQYIILAGPIQNYISITLWDFDFSTQLRKP	120		
Db	61	EILRQLETTISTIKYNSKOPVKNLTNTWTFPPQYIILAGPIQNYISITLWDFDFSTQLRKP	120		
Qy	121	AKYVYSQYNHTAKTITFR-PPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMFF	179		
Db	121	AKYVYSQYNHTAKTITFRPPPPCGRVPSMTCLSEMLNVSKRNDTGQCGNFTTFNPMFF	180		
Qy	180	NVPRWNTKLYVGTQKYNVDSQTIYFPLGLTALLRYAQRNCTHSFYLVNAMSRLNRPVKY	239		
Db	181	NVPRWNTKLYVGTQKYNVDSQTIYFPLGLTALLRYAQRNCTHSFYLVNAMSRLNRPVKY	240		
Qy	240	INGTKLNTMRKLKRQADVKQFQEKAKKTOSTTTPYFSYTTSAALNTVNTYITTA	299		
Db	241	INGTKLNTMRKLKRQADVKQFQEKAKKTOSTTTPYFSYTTSAALNTVNTYITTA	300		
Qy	300	ARRVSTSTIAYRPDSSPMKSI MATQLRDLATVWYTTLRVQRNPFCEPSNRNRTAVSEFMKN	359		
Db	301	ARRVSTSTIAYRPDSSPMKSI MATQLRDLATVWYTTLRVQRNPFCEPSNRNRTAVSEFMKN	359		

QY 360 THVLIRNETPTTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWD 419
Db 360 THVLIRNETPTTYIGTLDMSLYNETMFVENKTASDSNKTTPSPSMGFORTFIDPLWD 419
QY 420 YLDSLLFLDEIRNFSRLSPPTVNLTPPEHRRAYNLSSTNSLWWLQ 465
Db 420 YLDSLLFLDEIRNFSRLSPPTVNLTPPEHRRAYNLSSTNSLWWLQ 465

RESULT 2

US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706
; GENERAL INFORMATION:
; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoengendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaesenoiff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921

Query Match 23.5%; Score 575; DB 2; Length 15281;
Best Local Similarity 1.7%; Pred. No. 2.2e-06;
Matches 254; Conservative 78; Mismatches 128; Indels 14673; Gaps 82;

QY 1 MG----- 2
Db 1 MGAIGQDMAYDLNLPSSASSISNRYSEPVEQSPAQGLWFLHQLKLGASWDITPAAIR 60
QY 3 ----- 2
Db 61 LRGLHDIDALNAASRALTORHETLRTTFKEQDGVGVQVWVHAGLERGLRVDAASRDLAQ 120
QY 3 ----- 2
Db 121 LLAEEQTMKFDLESEPAWRVALLKVAEDHHILSIWVHHIISDRSLDIIQQELGELYTAA 180

QY 3 ----- 2
Db 181 SQKKSISACPLGPPIQIYRDLTTWQNOQVAEQERQIGYWIQLDNNTPAELLTBLPRP 240
QY 3 -----RKEMMV----- 8
Db 241 AIPSGETGKISFQIDGSHVHKLAPCRSQQVTAYAVLLAAFRVAHFLTGAEDATIGAPV 300
QY 9 -----RDVP----- 12
Db 301 ANRDRPELENVAPLATLQCMRVVLDDDTFESVLRLQIMSVMTTEAHANDVFFERIVSAL 360
QY 13 -----KMFVL----- 17
Db 361 LPGSTDTSRHPLVQLMFALHPAODTGRARWGFLAEETLQSAAPTRFDMEMHLFEGDDRFD 420
QY 18 ----- 17
Db 421 ANVLFTGLFADABAIRSVSVSIFREVLRRGISSEPAHVHVKTMPLTDGLAAIRDMGLLDIGTT 480
QY 18 ----- 17
Db 481 DYPREASVVDMEQVQVNLNPSATAVADASSRLSYSELDHKSQDLAAWLRRLRQLKPTLIG 540
QY 18 ----- 17
Db 541 VLSPPSCETWVSFLGILKAHLAYLPLDINVPLARIESILSAVDGHKLVLLGNSVNPQPKVD 600
QY 18 ----- 17
Db 601 VPDVELLRISDALNGSQVNLGAGKQATAPKATDILAVIFTSGSTGKPKGVMIHRGIVR 660
QY 18 ----- 17
Db 661 LVKGTNIISPAQAAVPTAHLANTAFDLSTWEIYTPILNGGTILVCIHESVTLDSKALEAVF 720
QY 18 -----ISISFL----- 23
Db 721 TKEGIRVAFAPALI KQCLADRPAPAGLDSLVAITGDRDRDALHAKSLVKHGVYNAYG 780
QY 24 ----- 23
Db 781 PTENSVVSTIYSVSEASPVPTGVVGRRAISNGAYVMDQDQLVSPGVMGELVVSQGLA 840
QY 24 ----- 23
Db 841 RGYTDSALDKNRPVVVQIDGESIRGYRTGDRARYSLKGGQLEFPGRMDQQVKIRGHRIP 900
QY 24 ----- 23
Db 901 AEVEHALLNSQVDRDAAVVIRRQEEEPAMIAFVTTQGTLPDLHVLNNGHVPDNGSK 960
QY 24 ----- 23
Db 961 NDQFAVHVESELRRRLQMLLPYSYMPARIIVLDHPLINPNKGVDRKALQSQAKTVQSKL 1020
QY 24 ----- 23
Db 1021 VSORVAPRNEIEAVLCEYSYVLGVEVGTDNPFDLGGHSLTAMKLAARISQRLDIQASV 1080
QY 24 ----- 23
Db 1081 ATVPEQPMADLAATTQRGSTLYSVIPTTEYTPVEQSPAQGLWFLHQLTASWYVM 1140
QY 24 ----- 23
Db 1141 LTVRLRGLHDVLDALGTALLALEKRHETLRTTFEERDGVGMQVHSSLMGELRLIDISEKS 1200
QY 24 ----- 23
Db 1201 GTAAHEALMKEQSTFDTLTREPGMRVALLKLADHHIFSIWVHHIVSDGWSLDRHQLG 1260
QY 24 ----- 23

Db 1261 LYSALRGDPLSRLEPLIOYRDFAVWQKDSQQKAAHQRLQLEYWTKQLADSTPAELLT 1320
QY 24 ----- 23
Db 1321 DFPRLSGKAGKVPVAIEGSLYDTLQVSRTHQVTSFAVLAAPRAAHFRLTGSNDAT 1380
QY 24 ----- 23
Db 1381 IGVSANRRPELENVIGFVNTQICIRITIDENDFESLVQRVRSITTAQNDQVPFEQ 1440
QY 24 ----- 23
Db 1441 VVSLMPSSRSDASRNPVLQMFALHGQDLFKIQLEGTEBEVIPTEVTRFDIEPHLYQ 1500
QY 24 ----- 23
Db 1501 GASKSGDIIPAADLPEAETIRGVSVFQEVLRGLQQPOTPIWMLPTDGIPELERMGL 1560
QY 24 ----- 23
Db 1561 LHMVKTDYPRNMSVDVFOQQVRLSABEATAVIDSSRMSYAELOQSDQVAAMLQRQLP 1620
QY 24 ----- 23
Db 1621 AETFVAVLAPRSCAVIALFGILKAGHAYLPDNNVPAARLAILAEVKGKLVLLGAGE 1680
QY 24 ----- 23
Db 1681 PSEGOPEVSIURIADATSPAGHASLRDCKSKPTAGSLAYVIFTSGSTGKPKGVMEHR 1740
QY 24 ----- 23
Db 1741 GVLRVKTNILSSLPAPQTFRMAHNSLAFDASIEVFTALLNGGSLVCIDRFTILDAQ 1800
QY 24 ----- 23
Db 1801 ALBALFLREHINIALPPALLKQCLTDAATIKSLDLLYVGGDRDLPDADAALAKLVKSE 1860
QY 24 ----- 23
Db 1861 VYNAYGTENTVMSTLYSIADTERFVNGVPIGRAVNSGVYVMDQNLVPLGVNGELVV 1920
QY 24 ----- 23
Db 1921 TGDGLARGYTNPALDSRFDVVIARGQLLRAVRTGDRARYRPKDQGVFFGRMDHQVKVR 1980
QY 24 ----- 28
Db 1981 GHRIELAEVEHALLSSAGVHDVAVVNSQBDNOGVENVAFITAQDNETLQEAQSNQVQE 2040
QY 29 ----- 28
Db 2041 WESHFETAYADITADQNTLGRDFTSWTSMYDGLIDKREMQEWLDDTMRFTLDGQAAG 2100
QY 29 ----- 28
Db 2101 HVLEIGTGTGWLFLNLQAGLKSIVGLEPSQSAVFVNKAAQTFPGLEGKAQVHVGTAMD 2160
QY 29 ----- 28
Db 2161 TGRLSALSPDLIVINSVAQYFPPSREVLAEVVEALVRIPGVRRIFPGDMRTYATHKDFLVA 2220
QY 29 ----- 28
Db 2221 RAVHTNGSKVTRSKVQOEVARLEBELELLVDPAFTSLKESLSEBIEHVLEILPKMKVN 2280
QY 29 ----- 28
Db 2281 NELSSYRYGAVLHRNHQNSRSIHKINAESWIDPASSQMDRQGLARLLKENKDAESIA 2340
QY 29 ----- 28

Db 2341 VFNIPYKTIIVERHIAKSLADDDHDDTHSSIDGVAMIISAAREKASQCPSLDVHDLVOLA 2400
QY 29 ----- 28
Db 2401 EDAGPRVEVSUARQSRONGALDVFFHHFQPTENESRALVDFPTDYKGOQARSLNRPLQR 2460
QY 29 ----- 28
Db 2461 VESRRIEAQVREQLQVLLPAYMIPARIVVLQNNPLNTSGKVDKRLTLRAKVTAARTPSS 2520
QY 29 ----- 36
Db 2521 ELVAPRDSIEAIIKCFKDVLGVEVGITDNFNVGSHSLATYKLAARLSQNLQAIAVKD 2580
QY 37 LYNRP ----- 41
Db 2581 IFDRPVIADLAATIQQDTTEHNPILPSTYGPVEQSFQAQRLWFLDQLNVGATWYLMPPA 2640
QY 42 ----- 41
Db 2641 VRLRGPLVVSALAAALLALEERHETLRTTFIEOGIGMQVIHPFAPKELRVIDVSGEES 2700
QY 42 ----- 41
Db 2701 TIQKILEKEQTTPFNLAASEPGFRLALLKTGEDEHILSTVWHHAISDQWSVDIFQOEIGQF 2760
QY 42 ----- 43
Db 2761 YSAILRGHDPLAQIAPLSIQYRDFATWQRQIFQVAEHRQLAYWTKQLADNKPALLTDF 2820
QY 44 ----- 49
Db 2821 KRPPMLSGRAGEIPVVDGLIYEKLQDFCRIRQVTAFTVLLAAFRAAHYRMTGTEDATIG 2880
QY 50 ----- 49
Db 2881 TPIANRRPELEGLIGFNTQCMRITVDVESFETLVHQVRETTLAHANQDVPEQIV 2940
QY 50 ----- 69
Db 2941 SNILPGSSDTSRNPLQMLFALHSQQNLGKVRLEGIEEBIISIAETTRPDIIEPHLYQEA 3000
QY 70 ----- 69
Db 3001 RLNGSIVYAADLFVPETIQSVITIFQGILOKGLGEPDMPVSMALDGLSLRSTGLLHP 3060
QY 70 ----- 69
Db 3061 QOTDYPCDASVVQIFKQVAVNPDVIAVRDESTRLSYADLDRKSDQVACWLSRRGIAPET 3120
QY 70 ----- 69
Db 3121 FVAILAPRSCETTIVAILGVLKANLAYLPDNNVNPASRLBAILSEVSGSMLVLVGAETPIP 3180
QY 70 ----- 69
Db 3181 EGMAEATIRITIELADAKTDINGLAASQTAASLAYVIFTSGSTRPKGVNVHREGIV 3240
QY 70 ----- 69
Db 3241 RLTKQNTITSKLPESPHMAHINLAFDASVWEVFTLLNGGTLVCIDYFTLLESTALEKV 3300
QY 70 ----- 69
Db 3301 PFDQVNVALLPPALLKQCLDNSPALVKTLISVLIYIGDRDLSDAANKARGLVQTOAFNAY 3360
QY 70 ----- 80
Db 3361 GPTENTVMSTIPIAEDDPINGVPIGHAVNSGAFVMDQNOQTTPGAMGELIVTGDGLA 3420
QY 81 ----- 80
Db 3421 RGYTSSLNTGRFINVDIDGEOVRYRTGDRVRYRPKDLQIEFFGRIDHQVKIRGHRIEP 3480

Qy	81	-----	130	-----	129
Db	3481	AEVEYALLSHDLVTDAVTHSQENODLEWVGFVAARVADVREDESSNQVQEWQTHFDSI	Db	4561	FHEQVSNPDSIALIHGSEKLSYAQLORESDRVARWLRHRSFSDTTLIAVLAPRSCETII
Qy	81	-----	130	-----	129
Db	3541	AYADITTIQQSLGRDFMSWTSMYDGLIKKQSQWELDDTMRSLDSQPPGHVLEVTG	Db	4621	AFLGILKANLAYLPDVKAPAARIDAIVSSLPGNKLILGANVTPPKLQEAADFPVIRD
Qy	81	-----	130	-----	129
Db	3601	TGMVFNLEGLGLOSYVLEPSPSATAFVNKAAGKSPGLEDRIRVEVGATDIDRLGDD	Qy	130	-----
Qy	81	-----	130	-----	129
Db	3661	LHAGLVVNSVAQYFSDYLAQLVRDLTKVPGVERIFFGDMRSHAINRDLVARAVHAL	Db	4681	TFTLTDTLQDGPTEIRPSAQSLAYAMFTSSGTGRPKGVMOVHRNIVRLVKNVNAKQ
Qy	86	-----	130	-----	129
Db	3721	GDKATKAEIQREVVRMESEDELLVDPAPFTSLTQVENIKHVEIILPKRMATNELSVYR	Qy	130	-----
Qy	86	-----	130	-----	129
Db	3781	YAAVLHVNDLAKPAHKVSPGAWVDFAATKMDRDALLRLRGTKISDHIAIANPNSKTIV	Db	4741	PAARIAHISNLAFDASSWEIYAPLLNGAIVCASYFTTIDPOALQETFOBEHIEIRGMLP
Qy	86	-----	130	-----	129
Db	3841	ERTICESYDLDGDAKSDNRVSWLSAARSNAVKNASLAIDLVDIAQAGPVEISCAR	Qy	130	-----
Qy	86	-----	130	-----	129
Db	3901	QMSQNGALDAVPHLGPSPQSSHVIDFLTDHQRPEEALTNHPLHRAQSRVERQIRER	Db	4801	PSLLKQCLVQAPDMISRLDILFAAGDRFSSVDALQAQRLVGSVFNAYGPTENTILSTIY
Qy	86	-----	130	-----	129
Db	3961	QNTFFQOYILA-----	Qy	130	-----
Qy	86	-----	130	-----	129
Db	4021	COBFSVDLGDGIMENFDLGGHSLMATKLAARISRLETHSVKVEIFDHPVCDLVI	Db	4981	GFVEDVAIVIRTPENQEPENVAFTAKGDSARBEZATTQIEGWEAHFEGGANIEEIE
Qy	86	-----	130	-----	129
Db	4081	VQGSAPHPDPIVSTKYTFVQSFQAGRLWFLDQNFCAWYMLPLAVLRGAMNVHALT	Qy	130	-----
Qy	111	-----	130	-----	129
Db	4141	AALLALERRHELLRTTFYEQNGVGMQVNPVVTETLRIIDLNGDGDYLPPLKKEQTAPF	Db	5041	SEALGYDFMGWTSVMYDGEIDKDEMRWINDTMRSLDGRKPGRVLEVTGTGIMFNLG
Qy	111	-----	130	-----	129
Db	4201	HLETPGWRVALLRPGDYILSVVMHHIISDGMSVDVLFQELGQFYSTAVKGDHPLSQT	Qy	130	-----
Qy	118	-----	130	-----	129
Db	4261	TPLPHYRDFALMOKKPTQSEHEROLQYVWEQVDSAPAEALLDLPSPSILSGQAGEMS	Db	5101	RSQGLERYIGLEPAPSAAEFVNNAKSPGLAGRAEVHVGTAAVCTLOGLTSDMAVINS
Qy	121	-----	130	-----	129
Db	4321	VTIEGALYKNLEBFCVRHVTSFVLLAALRAAHYRLTGSEDATIGTPIANRNPPELQI	Qy	130	-----
Qy	124	-----	130	-----	129
Db	4381	IGFFVNTQCIITVNEDETPELSVQVQVSTATAFAHQDVPEKIVSTLLPGSRDASRNP	Db	5161	VAQYFPTPEYLAETIKSLVQVPCMKIYILGDMRSHANNRDFAAARAAYSLADNASKDRVR
Qy	124	-----	130	-----	129
Db	4441	LVQLMAVHSQKNLGBLENASHVVPTEITRFDLEBFLHQDDKLEGSILYSTDLFE	Qy	130	-----
Qy	130	-----	130	-----	129
Db	4501	AVSVQSLLSVFOEILRLRGLNGDPVPISTLPLQDGIQVLDQGLLDVOKTEYPRDSSVDV	Db	5221	QKMELLEKEEBELLVDPAPFTALASQLODRIQHVETLPRKMKATNELSSRYAAVHLISD
			Qy	130	-----

Db	5639	EQRHEPLRTTFBEDGVGVQVQDHRPKOLRIIDLSTQPKDAYLVLKHEQTTFLDLATE	5698
Qy	182	PRWNTKL-----	188
Db	5699	PGWRVALIRLGBEBHILSIWMHHIISDGWSEVLFDEMHRFYSSALRQDPMQILPLPI	5758
Qy	189	-----	188
Db	5759	QYRDFAAWQKTERQVAEHQRLDYWTEHLADSTPAELLTDLPRPSILSGRANELPLTIEG	5818
Qy	189	-----	188
Db	5819	RLHDKLRAFRCVHOATFFVILLALRAAHYRLTGASDATLGTPIANRNPPELENMIGFFV	5878
Qy	189	-----	188
Db	5879	NTQCMRIAENDNFESLVRVRSTATSAFANQDVPFESIVSSLLPGSRDASRNPLVQVI	5938
Qy	189	-----	188
Db	5939	LAVHSQDLGKLTLEGLRDEAVDSASTRPDVEFHLPEHADRLSGSVLYAKELFKLRTIE	5998
Qy	189	-----	188
Db	5999	SVVSFLETLRALDQPLTPLAVLPLTDGVGEIASKGLLDVPRTPYPRDANIIVEVFOHV	6058
Qy	189	-----	188
Db	6059	RATPDALAVKDATSILTYAQDQSDRLAIWLSRRHMPETLVGLAPRSCETIIAMFGI	6118
Qy	189	-----	188
Db	6119	MKANLAYPLDINSPAARLSILSAVDGNKLVLLGSGVTAPEQENPEVEAVGIOELAGT	6178
Qy	189	-----	190
Db	6179	GLDKQSNARPATSILAYVIFTSGTGKPGVMVEHRSVTRLAKPSNVISKLPQGARVA	6238
Qy	191	-----	190
Db	6239	HLANIAFDASIWELATLLNGATLVCLDYHTVLDCTRLKEVPERESITVVTMLPALLKQC	6298
Qy	191	-----	203
Db	6299	VASIPETLAHLDLLTYGDRVGHDAMARSLVKIGMFGSGYPTENTVIS-TIYEVDADE	6357
Qy	204	-----	203
Db	6358	MFYNGVPIGKTNSGAYVMDRNQQLVPSGVVGGELVVTGDGLARGYTDPSLNKNRFIYT	6417
Qy	204	-----	203
Db	6418	VNGESIRAYKTGDRVRYRPHDLQIEFPFGMDQVKIRGHRIEPGEVESALLSHNSVDAA	6477
Qy	204	-----	203
Db	6478	VVICAPADQDSGAEMVAFVAARNTEDEDTQEEBAVDQVQGWETHFETAAYSEVKDIRQSE	6537
Qy	204	-----	210
Db	6538	VGNDPMGWTSMYDGEIDKTMHEWLNDRMTILDAREPHGVLEIGTGTGMVFNLAACP	6597
Qy	211	-----	210
Db	6598	GLOQYVGFEPKSAAPFNDAASFPALKDGRSIVHVGTATDINKAGPIQPRLVVINSVA	6657
Qy	211	-----	210
Db	6658	QYFPTPEYLFRVVEALVQIPSEVERIVFGDMRTNAINRDFVASRALHTLGEKANKELVRQM	6717
Qy	211	-----	210

Db	6718	IYELEANESELLTDPAFFTSLRTRLGEKIKHVEILLPKTKATNELSKYRYAAVLHVRGSR	6777
Qy	211	-----	210
Db	6778	EQSTIHQVSPNAWIDPAADGLDRQTLINLLKEHKDAGTVAIGNIPYSKTIVERFNKSL	6837
Qy	211	-----	220
Db	6838	EDMEEGQNSLPGSAWVAARMAOCSPLDAMDVKIEAQEAGYQVEVSWARQWSONGAL	6897
Qy	221	-----	220
Db	6898	DAIFHFEPKEGARTLIEFPTDYEGRNVTLTNRPLNSIQSRRLGTQIREKLTLLPPY	6957
Qy	221	-----	220
Db	6958	MIPSRIMVLQMPVNNNGIKDRKELVRAIVAPKPRSAATRVAPRNEIEAILRDEPEDVL	7017
Qy	221	-----	222
Db	7018	GTEVSVLDNFPDLGGHSLMATKLAARVSRRLDAHISIKOVFQDQVLADLAASIQRESAPH	7077
Qy	223	-----	222
Db	7078	EPQPQPYTGPAEQSPAQRLLWFLDQNLGATWYLMPLAIRRGQLRVAALSAAALFALER	7137
Qy	223	-----	222
Db	7138	RHETLRTTTEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYLEVLEVRREQTVFPDLSPEG	7197
Qy	223	-----	222
Db	7198	WRVCLVKTGEEDHVLIVMHHIYDGSVDIILRGELGQFYSAALRGQDPLLANPLIOY	7257
Qy	223	-----	222
Db	7258	RDFAAWQREAKQVEEHQRLGYWSKQLVDSTPAELLTDLPRPSILSGRAGSVDTVIEGSV	7317
Qy	223	-----	222
Db	7318	YGALQSFCTRSTVTFVLLTVFRIAHFRLTAVDDATIGTPIANRNRPELTLVGCFVNT	7377
Qy	223	-----	222
Db	7378	QCMRISIADDNDFEGLVRQVRNVATAAYANQDVPFERIVSALVPGSRNTRNPLVQLMFA	7437
Qy	223	-----	226
Db	7438	VOSVEDYDQVRLEGLESVMPGEASTRFDMFHLVPGDQKLTGSLVLYSSDLFEOGTIQNP	7497
Qy	227	-----	235
Db	7498	VDIFQECLRSVLDQPLTPISVLPFSNAIINLESLDLLEMTSDYPRDRTVVDLFRQAAI	7557
Qy	236	-----	235
Db	7558	CPDSIAVKDSSQLTYAQLEQSDRVAAWLHERHMPAESLVGLVSPRSCETIIAYFGIMK	7617
Qy	236	-----	238
Db	7618	ANLAYPLDVYAPDARLAAILDVTEGERLLLLGAGVPQPGIQIPRISTAYIAEALSHATT	7677
Qy	239	-----	238
Db	7678	VDVTSIQPSATSALAVIFTSGTGKPGVMIEHRCIVRLVRDTRNVNVPFESSALPVSH	7737
Qy	239	-----	238
Db	7738	FSNLAWDAATWEIYTAVLNGGTVCIDRDTMLDIAALNSTFRKENVRAAAFPTPAFLKQCL	7797
Qy	239	-----	239
Db	7798	AETPELVANLEILHTAGDRDLDPGDANLAKGTAKGGIFNVLGHNTENTAYSTFYFVVGEEF	7857

QY 240 ING----- 242
: ||
7858 VNGVPVGRGISHAYIIDRHQKLPAGVGMELITLTDGVARGYTDSALNKRFPVYIDIN 7917
QY 243 ----- 242
7918 GKSTWSYRTGDKARYRPRDQLEFFGRMDQWKIRGVRIEPEVELTLLDHKSVLAAATVV 7977
QY 243 ----- 242
7978 VRRPPGDPEMIAFITIDAEDDVOTHKAIYKHLOGILPAYMIPSHLVILDOMPVTNGKV 8037
QY 243 ----- 242
8038 DRKDLALRAQTVQKRRSTAAVRPRDEVAEVLCEBSNLEVEGITDGFDDLGGHSLLA 8097
QY 243 TKLK-----NT----- 248
|||
8098 TKLAARLSRQLNTRSVKDVDFQPIFADLADIIRRGSHRDPPIPATPYTGVEQSFQAGR 8157
QY 249 ----- 248
8158 LWFLEQLNUGASWYLPFAIMRGPLOTKALAVNALVHRHEALRTTFEDHGVGVQVI 8217
QY 249 -----MRKLRKQA----- 257
: |||
8218 QPKSSQDLRIIDLSDAVDDTAYLAALKREQTAFDLTSEPGWRVSLRLGLDDVILSIVM 8277
QY 258 -----PVKEQ----- 262
: |||
8278 HHIIISDGTVDLRLQELGQFYSAAIRGQEPLSQAKSLPIQYRDFAVMWQROENIQEQAQ 8337
QY 263 ----- 262
8338 LKYWSQQLADSTPCEFLTDLPRPSILSGEADAVPMVIDGTVYQLLTDPCRTHQVTSFVUL 8397
QY 263 -----FEKKAK 269
: |||
8398 LAAPRTAHYRLTGTLDATVGTPIANRNRPELEGLIGFVNTQCMWAISSETETESLVLQ 8457
QY 270 TQSTTTPYFS----- 279
: |||
8458 VRLTTTEAFANQDVPPEQIVSTLLPGSRDTSRNPVQVMPALQSQDGLRQLEGMTDEA 8517
QY 280 ----- 279
8518 LETPLSTRLLDLEVLHFOEVBKLSGLLYSTDLEFEVETIRGIVDVFLEILRRGLEQPKQRL 8577
QY 280 ----- 279
8578 MAMPITDGITKLRDQGLLTVAKPAYPRESSVIDLFRQVAAAAPDAIAVWDSSTLTLYADL 8637
QY 280 ----- 279
8638 DGQSNKLAHWLCORNMAPETLVAVFAPRSLCTIVAFLGVKANLAYPLDVNAPAARIEA 8697
QY 280 ----- 279
8698 ILSAVPGHKLVLVQAHGPELGLTMADELVLQIDEALASSSGDHEQIHASGPTATSLAVV 8757
QY 280 ----- 279
8758 MFTSGTGPKGVIMDHRSIIRLVKNSDVVATLPTFVRMANVSNLAFDISVQBIYITALIN 8817
QY 280 ----- 279
8818 GGTIVCLDYLTLLDSKILYNVFEAQVNAAMFTPVLLKQCLGNMPAIIISRLSVLFNVGDR 8877
QY 280 ----- 279
8878 LDAHDAVAASGLIQDAVINAYNGPTENGQSTMYKVDVNEPFFVNGVFIGRSITNSGAYVMD 8937

QY 280 -----YTTSA----- 284
: |||
8938 GNOQLVSPGVMEIIVTGDGLARGYITDSALDEDRFVHVITIDGEENIKAYRTGDRVYRPK 8997
QY 285 ----- 284
8998 DFEIEFFGRMDQOVKIRGHRIEPAEVEHALLGHDLVHDAVAVLKRPANQPEMIAFITSQ 9057
QY 285 ----- 284
9058 EDETIEQHSNKOVOQWGEHFDVSRVADIKOLDTSTFGHDFLGWTSYMGVDVIPVNEMKE 9117
QY 285 ----- 284
9118 WLDETTASLLDNRP PGHILEIGAGTGMILSNLKVLDGLQKYVGLDPAPSAALFVNEAVKS 9177
QY 285 ----- 284
9178 LPFLAGKARVLVGTALDIGSLDKNEIQPELVINSVAQYFPTSEVYLKVVKAIVEVPSVK 9237
QY 285 -----ALN----- 287
: |||
9238 RVFFGDIRSQALNRDPLAARAVRALGDNASKEQIREKIAELESEEBELLVDPAFFVSLRS 9297
QY 288 ----- 287
9298 QLPNIKHEVLPKLMKATNELSSRYAAVLHISHNEEBEQLLIQDIDPTAWVDFAAATQKDS 9357
QY 288 -----VTNTVYSITTAARRV----- 303
: |||
9358 QGLRNLQOGRDVMIAVGNIPYSKTI VERHIMNSLDQHVNSLDGTSWISDARSAAATC 9417
QY 304 ----- 303
9418 TSFDAPALTOLAKEGFRVELSWARQSQNGALDAVHRLATDANCERSRVLHPPTDHO 9477
QY 304 ----- 303
9478 GRQLRTLNRPLQRAQSRRIESQVFEALQALPAYMIPSIIVLPQMPNTNANGKVDKQOL 9537
QY 304 ----- 303
9538 ARRAQVAKRKAVSARVAPRNDTEIVLCEYADILGTEVIGITDNFDFMGHSLMATKLAA 9597
QY 304 ----- 303
9598 RLSRRLDTRVTYKVEFDKPEVLADLAASIEQGSTPHLPIASSVYSGPVEQSYAQGRLWFLD 9657
QY 304 ----- 303
9658 QFNLANATWYHMSLAMRLLGLPLNMDALDVALRALEQRHETLRTTPEAOKDQIGVQVVEAGM 9717
QY 304 ----- 303
9718 KRLKVLDSLKNEKEHMAVLENEQMRPFTLASEPGWKHGLARLGLPTEYILSLVMMHMFSD 9777
QY 304 ----- 303
9778 GWSVDILRQELQGFYSAAALRGDRDPLSQVKPLPIQYRDFAAWQKEAAQVAEHERQLAYWEN 9837
QY 304 ----- 303
9838 QLADSTPGELLTDFPRPQFLSGKAGVIPVTIEGPVYEKLLKFSKERQVTLFVLLTAFRA 9897
QY 304 -----STST 307
: |||
9898 THFRLTGAEDATIGTPIANRNRPELEHIIIGFPVNTQCMRLLLDGTGSTFESLVQHVRSVAT 9957
QY 308 IAY----- 310
: |||
9958 DAYSNOQDIPFERIVSALLPGSRDASRSLIQLMFALHSQPDIGNITLLEGLERLPTSPA 10017
QY 311 ----- 310

Db	10018	TRFDMFHLFQEPNKLGSILFADELFPQPETINSVVTVFOEILRRGLDQPVSI	10077
QY	311	-----	310
Db	10078	DGLIDLEKLELLESSNPPRDYVVDVFPQVAAANPAPVVDSETSMYS	10137
QY	311	-----	314
Db	10138	IAAWLHAQGLRPESLTCVMAPRSFETIVSLFGLKAGAYPLDVPNSPAARIQ	10197
QY	315	-----	314
Db	10198	GKRLVLLGSGIDMPQSDRMDVETARIQDILNTTKVRSDDMSRPSATS	10257
QY	315	-----	314
Db	10258	RPKGVMEHNRILRLVKQSNVTSQPLQDLRMAHISNLPADASIWEIFT	10317
QY	315	-----	318
Db	10318	YFTLLDQALRTTFEKARVNATLFPALLKECLNHAPTLPEDLKVLVIG	10377
QY	319	-----	318
Db	10378	IQALVRGTVNAYGPTENTVMSTIYRLTDCESYANGVPIGNVSSGAYIM	10437
QY	319	-----	318
Db	10438	GVMGELVSGDLGARGYTNSTLNADRPDIVINDQKARYTGDRTRYR	10497
QY	319	-----	318
Db	10498	MDQQVKIRGHRVEPAEVEQAMLGNKALHDAVVVQAVDQGETEMIGF	10557
QY	319	-----	318
Db	10558	EITNQVQWEDHFEATAYAGIEAIDQATLGRDFTSWTMYNGNLIDKA	10617
QY	319	-----	318
Db	10618	LLDKEDARPCAEIGTGTGMVLFNLPRKNDGLSVGIEPSRSALFVDK	10677
QY	319	-----	318
Db	10678	QILVGTAEEDIKLVKDFHPDVVVINSVAQYPPSRSLVQIASELIHMT	10737
QY	319	-----	318
Db	10738	ATNRDFLVRALYTLGDKATKQIOEVARLENEDELLVDPAPFTSLTS	10797
QY	319	-----	318
Db	10798	ILPKRMTSNELSSRYAAVLHICRDEGRNRYGRRVHSVEENAWIDP	10857
QY	319	-----	318
Db	10858	MLDERDAKTVAGNIPHSNTINERHPTTSLDTEGEGIAQDSLDSAW	10917
QY	319	-----	318
Db	10918	CLSVTELVEIGQAAGFRVEVSWARQSQHGALDVFHHLDDRVGRVLI	10977
QY	319	-----	318
Db	10978	STGLTSRPLQRIQNRFPESQIREQLOTLPPYVPSRIVVLRMPLNAN	11037
QY	319	-----	325
Db	11038	ARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGVMDNPFELG	11097
QY	326	-----	325

Db	11098	RRLDTRVSVKDI PNQPILODLADVQTSAPHEAIPSTPYSGPVEQSF	11157
QY	326	-----	325
Db	11158	LNASWTHMPLASRLRGLRIEALQSALATIIEARHESLRTTPEEQD	11217
QY	326	-----	325
Db	11218	RIIDVSGTBEDAYLAALKQEQDAAFDLTABPGWRVALLRLGPD	11277
QY	326	-----	331
Db	11278	DILRQELGQLYSNASSQPAPLPFIQVRDPALWQKDSQIAEHQK	11337
QY	332	-----	339
Db	11338	LADFTRPKALSGDADVIPIEIDDQVYONLRSFCRARHVT	11397
QY	340	-----	339
Db	11398	ATIGSPIANRNPELEGLIGCFVNTQCLRI PVKSEDTPTLVKQARE	11457
QY	340	-----	343
Db	11458	ERIVSSWVASSRDTSRNPLVQVMFAVHSQHDIGNIRLEGVEGK	11517
QY	344	-----	343
Db	11518	EDQGLGNNVFKDLFESETIRSVAVFQETLRRGLANPHANLAT	11577
QY	344	-----	343
Db	11578	LQVNOPDYPRDASVIDVFREQVASIPKSIADASSQLTYTELDER	11637
QY	344	-----	345
Db	11638	PEELGVGLAPRSCETIIAFLGIKANLAYPLDVNPAGRIETIL	11697
QY	346	-----	345
Db	11698	QAVKLHANSVRTRISDALVESGSPTEELSTRPTAQSLAYVFT	11757
QY	346	-----	345
Db	11758	GITRLVNSNVAKOPAAAAIAHLSNIAFDASSWEIYAPLLNGT	11817
QY	346	-----	345
Db	11818	EAVFKQHHIRGAMLPALLKQCLVSAPTMISSLEILFAAGDR	11877
QY	346	-----	345
Db	11878	NAYGPTENTVLSITHIGENAEAFSGVPIGNVNSGAFVMDQ	11937
QY	346	-----	345
Db	11938	DGLARGYTDKLRVDRFIYITLDGNRVRAYRTGDRVHRPKDQ	11997
QY	346	-----	345
Db	11998	RIEPAEVEQALARDPAISDAVITQLTDEEBELVAFPSLKNAN	12057
QY	346	-----	353
Db	12058	GDEQHALLMENKIRHNLQALLPTYMIPSRRIHVDQLPVNANG	12117
QY	354	-----	356
Db	12118	SVSTVAPRNDIETIICKEFADILSVRIGITDNFDLGGHSLI	12177
QY	357	-----	356
Db	12178	VRDVFDPVVGQLAASIQQGSTPHEAIPALSHSGPVQSQF	12237

QY	357	-----MKNTHVLIR	365	QY	414	-----	413
Db	12238	PGVRLRGRLVDALQALRALBERHELLRTTFEBQDGVGMQIVHSPMRDICVVDISGA	12297	Db	13318	GSNASKAMVRQOVAKLEDDEEELLVDPAFTLSLSDQFPDEIKHVEILPKMATNELSSY	13377
QY	366	NE-----	367	QY	414	-----	413
Db	12298	NEDLAKKEBOQAPFNLSTEVAMRVALFKAGENHHILSIVMHHIISDGWSVDIFQQLAQ	12357	Db	13378	RYAAVIHVGGHQMPNGEDEDKQWAVKDINPKAWVDFAGTMDRQALLQLLQDRQRGDDVV	13437
QY	368	-----	367	QY	414	-----	413
Db	12358	FYSVAVRGHDPLSQVKPLPIHYRFAVWQODQVAVHESOLOYWIBQLADSTPAEILSD	12417	Db	13438	AVSNIPIYSKTIHERHLSQSLDDDDCTSAVDGTAMISRQSRRAKECPALSVADLIEIGKG	13497
QY	368	-----	367	QY	414	-----	413
Db	12418	FNRPEVLGSEAGTPIVIEDEVYEKLSIFCRNHQVTSFVLLAARVAHYRLTGAEDATI	12477	Db	13498	IGFEVASWARQHSQRGGLDVAFHRPEPPRHSCHVMFRPTEHKGSRSSSLTNRPJLHLQ	13557
QY	368	-----	367	QY	414	-----	413
Db	12478	GTPIANRNRPELDLIGFVNTQCMRIALEBHDNPLSVVRVRSTAASAFENQDVPPERL	12537	Db	13558	SRLEAKVRERLQSLPPYMIPSRITLLDQMLTNSNGKYDRKKLARQARVIPSAASTLD	13617
QY	368	-----	371	QY	414	-----	413
Db	12538	VSALLPGSRASRNPVLQMFVHVSQRNLKQLQLEGLEGFTPYATTATRFDFVPHLFEQD	12597	Db	13618	FVAPRTEIEVILCEEFTDLLGVKVGITDNFFELGHHSLATKLSARLSRRLDAGITVKQV	13677
QY	372	-----	371	QY	414	-----	418
Db	12598	KGLAGNVFAADLPEAATIRSVVEFHEILLRGLDQPDIALSTMPLDGLAALNSRLPA	12657	Db	13678	FDQPVLDLAASILQOSSRHSRISPSLYEGPVEQSPAQGRMLWFLQDFNIDALWYLIPFAL	13737
QY	372	-----	377	QY	419	-----	418
Db	12658	VEDIEPDFATEASVVDVFQVWVANDALAVDTSTKLTAAELDQQSDHVAAMLSKQKLP	12717	Db	13738	RMRGPLQVDALAAALVALERHESLRTTPEERDGVGIQVQVPLRTTKDIRIIDVSGMRDD	13797
QY	378	-----	377	QY	419	-----	418
Db	12718	AESIVVVLAPRSSETIVACIGILKANLAYLPMSNVPEARQALISEIPGEKFFVLGAGV	12777	Db	13798	DAYLEPQKEQOQTPFDLASEPGMRVALLKLGKDDHILSIVMHHIISDGWSTEVLORELQ	13857
QY	378	-----	377	QY	419	-----	419
Db	12778	PIPNKTDVVRMVFISDIVASKTDKSYSPGTRPSASLAYVIFTSGTGRPKGMVVEHRG	12837	Db	13858	FYLAASKGAPLSQVAPLPIQYRDFAVMQRQEEVAESQRQLDYWKQKQADSSPAELLAD	13917
QY	378	-----	377	QY	420	Y-----	420
Db	12838	VISLVQNASRIPOSRLMAHVSNLAFDASVWEIPTLLNGGTLFCISYFTVLDSKALSAA	12897	Db	13918	YTRPNVLGSEAGSVFVINDSVYKLSVFCRSRQVTTFTTLAAPRAAAHYRMTGSDDATI	13977
QY	378	-----	384	QY	421	-----	420
Db	12898	FSDHRINITLLPALLKQCLADAPSVLSSLESYIGGDRLDGADATKVVDLVKQKAYNAY	12957	Db	13978	GTPIANRNRPELENLIGCFVNTQCMRITIGDDETFESLVQVVRSTTATAFENQDVFFERI	14037
QY	385	-----	384	QY	421	-----	426
Db	12958	GPTENSVMSTIYTIETHETFANGVPIGTSGLPKSKAYIMDQDQQLVPAGVMGELVWAGDGL	13017	Db	14038	VSTLSAGSRDTSRNPVLQLLFAVHSQQGLGRIQLDGVVDEPVLSTVSTFRDLETHAFQEA	14097
QY	385	-----	385	QY	427	-----	430
Db	13018	ARGYDPSLNTGTRPIHITIDGKQVAYRTGDRVVRPRDYQIEFFGRLLDQOQIKIRGHRIE	13077	Db	14098	DLRLNGSVMFATDLFPQETIQGFVAVVEVLQGLEQPOSPATMPLAEGIAQLRDAGALQ	14157
QY	386	TMFVENKTASDS-----	397	QY	431	-----	435
Db	13078	PAEVEQALLSDSSINDAVVVAQNKEGLEMGVITTTAAQSVDKKEASNKVQEWEAHPDS	13137	Db	14158	MPKSDYPRNASLVDVFOQQAASPTVAVTOSTSKLTYAELDRLSQQAASVYLRQQLPAE	14217
QY	398	-----	397	QY	436	-----	437
Db	13138	TAYANIGGIDRDALGDFLSWTSMYDGSILPREMOEWLNDTMRSLDDNQPPFKVLEIGT	13197	Db	14218	TWAVLAPRSCETIIAFLAILKANLAYMPLDVNTPSARMEAISSVPGRRLLVSGVVRH	14277
QY	398	-----	407	QY	438	-----	437
Db	13198	GTGWLFNLGKVEGLQSYAGLEPSRSVTAWVKAETETPFSLAGSARVHVGTAEDISSIDG	13257	Db	14278	ADINVPAKTMJLISDVTGTDAIGTPBPLVVRPSATSLAYVIFTSGTGKPKGMVVEHRA	14337
QY	408	-----	413	QY	438	-----	437
Db	13258	LRSDLVVINSVAQYFSPREYLAELTANLIRLPQVKRIIFPGDMRTVATNKNDFLVARAVHTL	13317	QY	438	-----	437

Db 14398 ETEREQVRAAIPFPALLROCLVNMPDAIGMLEAVYVAGDRFHSRDARATQALAGPRVYN 14457
Qy 438 ---PT---YVN---LTPP---446
Db 14458 AYGTENAILSTIYNDIKHDPYNGVPIGSAVNSGAYVMDRQQLPFGVMGELVVTGE 14517
Qy 447 ---446
Db 14518 GVARGYTDASLDTDRFVTTIDGQRQAVRTGDRVYRPGFQIEFFGRLDQOAKIRGHR 14577
Qy 447 ---EH---448
Db 14578 VELGEVEHALLSENSVTDAAVLRTWEEEDPOLVAFTVTDHEYRSGSNEEDPYATQAA 14637
Qy 449 ---RAVNLSTGNS---459
Db 14638 GDMRKELRSLLPYMYVPSRVITILQMLNANGVKDRKDLARRAQMTPTASSSGPVHVAPR 14697
Qy 460 ---459
Db 14698 NETEAAICDEFETILGVKVGITDNFFELGHSLLATKLAARLSRRMGLRISVKDLFDDPV 14757
Qy 460 ---459
Db 14758 PVSAGKLEQQQFSGEDESSTVGVFPQLLPAEMSREIIQRDVVPQIENGHSTPLDMYP 14817
Qy 460 ---459
Db 14818 ATQTOIFFLHDKATGHPATPPLSLOFPETADCRRLASACAALVQHFDFRTVFSRGR 14877
Qy 460 ---459
Db 14878 FYQVLAHLDPVVEVETIEQELDEVALHEADKQQLRLGRAMLRIALKRGAKRLV 14937
Qy 460 ---LW---461
Db 14938 LRMSHLYDGLSLEHIVNALHALYDKHLAQAPKFGLYMHMASRAEGYNFRSLQGS 14997
Qy 462 ---W---462
Db 14998 SMTSLKRSVGALEAMTPSAGTWTQTSKIRIPPAALKNGITQATLPTAAVSLLLAKHTKST 15057
Qy 463 ---462
Db 15058 DVVFGRWVSGRODLSINCQDIVGCINEVPVRVIDEGDDMGGLLRAIQDQVTSRPHET 15117
Qy 463 ---W 463
Db 15118 LGLQEVKENC TDW 15130

RESULT 3
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 21.1%; Score 516; DB 3; Length 10182;
Best Local Similarity 2.2%; Pred. No. 7.1e-05;
Matches 225; Conservative 90; Mismatches 129; Indels 9652; Gaps 69;
Qy 2 GRKEMVRDVPKMFVLISIFLLVSPINCKVMS-
Db 9 GRN-----ICSPLLSKCMYSLSKSLTLKTYNFQITSNKKEKTSRI 49
Qy 35 -----34
Db 50 GVAIALNRRDKLQKPSIRKYAIGTSTVIATLVPMTNTHASADELNQNKLIKQLNQT 109
Qy 35 -----34
Db 110 DDDSNTHSQEIENNKQSSGKTESLRSSSTQANARLSQDKOTNETSQQLPTNVSD 169
Qy 35 -----34
Db 170 SINQSHSEANMNEPLKVDNSTMQAHSKIVSDSGNASENKHHKLTENVLAESRASKNDK 229
Qy 35 -----34
Db 230 EXENLQEKDKSQOVHPPLDKNALQAFDASYHNYRMI DRDRADATEYQVKVSTFDYVNDL 289
Qy 35 -----KAL-----37
Db 290 LGNNQNPSEQLVSAVQOLEKALEARTLPQOSTTEKGRRRSTRSVVENRSSRSDYLAR 349
Qy 38 -----YNR-----41
Db 350 TBYVVS KDDDDSGPPGTFPHASNRWPYNLPRSRNLRASDVQGNAYITTKRLKDGQYQW 409
Qy 42 -----WR 43
Db 410 DILFNSNHKGHEMYWYFGLPSDQTPGPTTIINRDSSTSTGCVGFGSGAPLPQFWR 469
Qy 44 ---GLVLSKI-----50
Db 470 SAGAINSSVANDFKHGSATYAFYDGVNNSDFARGGELYDFREGATQTKYKYGDEFAL 529
Qy 51 -----GKYK-----LDQL 58
Db 530 LNSEKPDQIRGLDTIYFSGKGDVSYRISFKTQGTARLYAAGARSGEYKQATYNNQL 589
Qy 59 KLE-----ILRQLETTIS-----71
Db 590 YVEPYKNYRNVQSNVQVKNRTLHLKRTIRQFDTLQRTDVPILDGSGSIDSVDPL 649
Qy 72 -----71
Db 650 SYVKNVGTGVLGIYPSYLPYNOERWQGANMAYQIEELFSQENLQNAARSGRPIQLVG 709
Qy 72 -----TKYNVSK-----78
Db 710 FDVEDSHHNPETLLPVNLYVKPELKHTIELYHDNEKQDRKFSVSKRAGHGVQVMGSL 769
Qy 79 -----QPVKN-----LTMN-----87
Db 770 HNTVSGSILPYQOEIRIKLTSNEPIKQSEWSITGYPTNTLQNAVGRTNNAEKNLALVG 829
Qy 88 -----TEPP-----91
Db 830 HIDPGNYFITVKFGDKVEQFEIRSKPTPRIITANELRGNPHHKEIRVTDIPNDTTAK 889
Qy 92 -----91
Db 890 IKLVMGTDGDHDPINPYTPVNTVVAEYHNDPCKNGVLTFRSSDYLDKDLPLSGEL 949
Qy 92 -----QY-----94
Db 950 KAIVYNYQVQSNFNSVPFSSDTPPTTINEPAGLVHVKYRGDHEVITLPTVDNTGSGSL 1009

QY 95 ----- 94
Db 1010 RDVNVLPQGWTKTFTINPNNTTEGLKLIGNIPSEAYNTTYHFNITATDNGSNTTNP 1069
QY 95 ----- 94
Db 1070 KTFILNVGKLADLPNVLGRDQLQVLWDPSSLSNSEREBVKRISSEANANIRSYLLQNN 1129
QY 95 -ILAG----- 98
Db 1130 PILAGVNGDVTFYRDGSDVDIDAENVITYEPERKSIIPSENGNTNKEAVITIARGQNT 1189
QY 99 ----- 98
Db 1190 IGBNLRKYFSLNSGDLPNRDFTSISAIGSLPSSSEISRLNVCNRYRNVAKNAYHKTQQ 1249
QY 99 ----- 98
Db 1250 ELNLKLIVEVNAPTGNRRVRYSTYLTNDTEINKIKQAFKAANSGLNLNDNDITVSNRF 1309
QY 99 ----- 98
Db 1310 DHRNVSVTVTRKGDLLIKEFPSSNLNMNPLRWNIIRDDYTIISWTSKIQGRNTDGLW 1369
QY 99 ----- 98
Db 1370 SPDHKSIIYKDATLGRQINTNDVLTLLQATAKSNLRSNINSNEKQLAERGSNGYSKSI 1429
QY 99 -----PTQ----- 101
Db 1430 IRDDGSEYLLNSNPICQLDLVEPDNGYGRQVSHSNVIYNEKNSIVNGQVPEANGASA 1489
QY 102 ----- 101
Db 1490 FNIDKVVKANAMGIMGVIYKQALYLPSPKGYIEKLGQNLSTWNVNIVYFVPSDKV 1549
QY 102 NVSITY----- 107
Db 1550 NPSITVGNDDHHTVYSGETFKNITINVNDVGLMTVASTSDSAITMTNRNNNELVQAPNVT 1609
QY 108 ----- 107
Db 1610 NSINKIVKVKATDKSGNESIVSFTVNIKPLNEKYRITTSNSSQTPVRISNIQNNANLSIE 1669
QY 108 ----- 107
Db 1670 DQNRVKSLSMTKILGRNVYNESNNDVRSQVSVKUNRSNNATVNVTTTFSDCGTTNTIT 1729
QY 108 ----- 107
Db 1730 VPKVHVLEVPTRTTVRGQPPTGKGTSPNDFFSLRTGPPVDARIIVWVNNQGPDISN 1789
QY 108 -----LWFDYSTQLRPAKYVYSO-----YNHT----- 131
Db 1790 QIGRDLTLHAEIFDGETPIRKDTTKLSQSTPKQIYETTINGRPNSSGDYAPGNFVOA 1849
QY 132 -----AKTIF-----RPP----- 140
Db 1850 VNQYWPEHMDFRWAGSGTSPSSNAGSFTKVTIVYQNGQTEENVNVLFKVKENKPVDSN 1909
QY 141 -----PCGRV 145
Db 1910 SVISKQLNGQQLIVRNPQNAQVTLYQSGNVTIPNTNTTIDSNGIATVTIQOTLPTGNI 1969
QY 146 PSMTCLSEMLNVSKRNDTG-----EQCGG-N 170
Db 1970 TAKTSMNNVTYTKQSSGLASNTTDDISVFSNSQVNVNVTAGMQANDGIIKIKGTNYN 2029
QY 171 FTTFNPMFFNVP-----RWN----- 185
Db 2030 FNDNSFISNIPAHSTLWNBEPSWKNNGITTKTIVTLPNHQGRTRVDIPITIIPTV 2089
QY 186 -----TKLY----- 189

Db 2090 TAKNPVRDQGRNLNTGTDVYNYIIPENNNRLGGTASWKDNRPQDKNIAGVQNLIALVNY 2149
QY 190 ----- 189
Db 2150 PGISTPLEVPVKVWVYNFDFTPQIYKIQVGDFPKGTWAGYKXHLNGBGLPIDGWKFYW 2209
QY 190 ----- 189
Db 2210 NQOSTGTTSDQWQSLAYTRTPFVKTGTYDVVNPSNWGVWQTSQSAKEIVTNAPNQPTIT 2269
QY 190 ----- 189
Db 2270 QSKTGDVTVTPGAVRNILISGTNDYIQASADKIVINKNGNKLTFVKNDGRWTVETGSP 2329
QY 190 -----VGPT----- 193
Db 2330 DINGIGPTNNGTAISLSRLAVRPGDSIEAIAIEGSGETISTATSIEIYVKAPQEQVAT 2389
QY 194 ----- 193
Db 2390 HTYDNGTDFILPDNSRNSLNPTERVEINYTEKLNNGNETQKSFITIKNNNGKWTINNKPNY 2449
QY 194 -KVNDSQTIYFL----- 205
Db 2450 VEFNQDNGKVFVSANTIKPNSQITITPKAGQGNENTNPTVIPAQAQHTLTINEIVKEQG 2509
QY 206 ----- 205
Db 2510 QNVTNDDINNAVOPVNKNRVAIKQGNALPTNLAGSTSHIPVVIYYSOGSSBEATVTRT 2569
QY 206 ----- 205
Db 2570 KVNKTELINARRLDEEISKENKTPSSIRNFQAMNRAQSQINTAKSDADQVIGTEFATP 2629
QY 206 -----GLT----- 208
Db 2630 QOVNSALSQVQAQNKINEAKALLQNKADNSQLVRAKEQLQOSIQPAASTDGMQTDSTRN 2689
QY 209 ----- 208
Db 2690 YKNKQAAEQATQHANSVINNGDATSQINDAKNTVEQAQRDYVEAKSNLRADKSQLQA 2749
QY 209 ----- 208
Db 2750 YDTLNRDVLTDKKPASVRRYNEAISNIRKELDTAKADASTLRNTNPSVEQVRDALNKI 2809
QY 209 -----ALL----- 211
Db 2810 NTVQPKVNOQIALLOPKENNELVQAKKRLQDAVNDIPOTQMTQQTINNYNDKQREABR 2869
QY 212 -LRYAQR----- 217
Db 2870 ALTSQRVIDNGDATTOEITSEKSKVEQAMQALTNAKSNLRADKQELQATYAKLIENTVST 2929
QY 218 ----- 217
Db 2930 NGKKPASIRQYETAKARIQNIQINDAKNEAERILGNDNPQVSVQVQALNKIKAIQPKLTEA 2989
QY 218 -----NCHTSFYLVMASR----- 231
Db 2990 INMLQKENNTELVNAKNRLNENAVNDTDPHGTQETINNYNAKKREAQNEIQANMIIN 3049
QY 232 ----- 231
Db 3050 NGDATAQDISSEKSKVEQVLQALQNAKNDLRADKRELQATYAKLIONVNTNGKKPSSIQN 3109
QY 232 ----- 231
Db 3110 YKSARRNIENQYNTAKNEAHNVLENTPTVNAVEDALRKINAIQPEVTKAINILOQKEON 3169
QY 232 ----- 231

Db	3170	SELVRAKEKLDQAINSPQLNGMTQESINNYTTTKRREAQNIASSADTIINNGDASIESQIT	3222
Qy	232	NLFRVPKYING-----TYLKNTWRKLKR	254
Db	3230	ENKIRVEEATNALNEAKQHLTADTTLSLKTKEVRKLSRRGDTNNKKPSSVSAVNTTIHSLQS	3289
Qy	255	-----	254
Db	3290	EITQENRANTIINKPIRSVEEVNHALHEVNQLNQLRDTLTINLLQPLANKESLKEARNRL	3349
Qy	255	-----KQAPVKEQ	262
Db	3350	ESKINETVQTDGWTQOSVENYKQAKIKAQNESSIAQTLINNGDASDOEVSTEIEKLNQKL	3409
Qy	263	-----FEKK-	266
Db	3410	SELTNSINHLTVNKEPLETAKNVLQANIDQKPSDTGMDTQOSVQSYERKLBQAKDKINSIN	3469
Qy	267	-----	266
Db	3470	NVLANPVDNAIRTNKVETEQINNELTQAKQGLTVDRKQPLINAKTALQSLDNQPSTTGM	3529
Qy	267	-----AKK-	269
Db	3530	TEATIQYNNAKQKAEQVIQNAWKIENAPQSVQOVSDSEKSVKEQALSELNNAKSALRAD	3589
Qy	270	-----	269
Db	3590	KQELQAYANQLIQPTDLNNKPASITAYNQRYQOFSNELNSTKNTDRILLKEQNPVSADV	3649
Qy	270	-----TQSTTTPYFS	279
Db	3650	NNALNKVREVQKLNARALLQNKEDNSALVRAKEQLQAAVDQVPSTEGMTQOTKDDYNS	3709
Qy	280	-----YTT-	282
Db	3710	KQAAQOEISKAQOVIDNGDATTOOISNAKTNVERALEALNNAKTGLURAKBELQNAYNQ	3769
Qy	283	-----SAALN-	287
Db	3770	LTQNDTSGKTPASIRKYNEAKSPIQTQIDSAKNEANSILTNDNPQVSQVTALNLKIKAV	3829
Qy	288	-----	287
Db	3830	QPELDKAIAMLNKKNENNALVQAKQQLQOIINEVDPTQGMTDTANNYKSKREAEDEIQ	3889
Qy	288	-----	287
Db	3890	KAQOIINNGDATEOOITNETNRVNQAINAKAKNDLRADKSLQLENAYNQLIQNVDTNGK	3949
Qy	288	-----	287
Db	3950	KPASIOQYQARQAIETQYNNAKSEAHQILENSNPVSNEVAQALQKVEAVQLKVNDIAHI	4009
Qy	288	-----	287
Db	4010	LQNKENNSALVTAKNVLQOSVNDQPLTTGMTQDSINNVEAKRNEAQSAIRNABAVINNGD	4069
Qy	288	-----	287
Db	4070	ATAKQISDEKSQVBQALHLNDAKQQLTADTTTELQTAVQQLNRRGDTNNKKPRISINAYNK	4129
Qy	288	-----	287
Db	4130	AIQSLQITSAKONANAVIQPRTVQEVNNALQQVNQLNQQLTEAINQLQPLSNNDAL	4189
Qy	288	-----	287
Db	4190	KAARLNLNKINQTVQTDGMTQOOSIEAYQNAKRVAQNESNTALALINNGDADEQOITTET	4249
Qy	288	-----VTTNVTYSI-	296
Db	4250	DRVNOOQTTNLTOAINGLTVNKEPLETAKTALONNIDOVPSDTGMDTQOSVANYNOKLOIAK	4309

Qy	297	-----	296
Db	4310	NEINTINNVLANNPDVNAIKTNKAEERIISNDLTOAKNNLQVDTQPLEKIKRQLQDEIDQ	4369
Qy	297	-----	296
Db	4370	GTNTDGTQDSVDNYNDSLAAIIEKGVKNLKRNPTEVOVKESVANAAQVTDLQNA	4429
Qy	297	-----	296
Db	4430	TSLVPDKTQLEAKNRLensiNQOTDGTQDSLNNYNDKLAKARQNLKISKVLGGQP	4489
Qy	297	-----	296
Db	4490	TVABIRQNTDEANAHKALDARSQTLNRPYINHHNESHLLNNAKQNFKAQVNSAPN	4549
Qy	297	-----	296
Db	4550	HNTLETKNKADTLNQSMWALSADIYENQKQENYLDASNKKRQDYDANVNAAGILN	4609
Qy	297	-----TTAARRVSTST	307
Db	4610	QTQSPTWSADVIDQABEDVKRTKTALDGNQRLVAKQALNHLNTLNDLDAQRTLDT	4669
Qy	308	IAYRP-----	312
Db	4670	INHSFNINSVNOAKEKANTVNTAMTQLKQTIANYYDDELHDGNYINADKKDAYNNAVNN	4729
Qy	313	-----	312
Db	4730	AKQLINOSDAQOLDPAEINKVTRVNTTKNDLNGNDKLAEAKRDANTTIDGLTYLNEA	4789
Qy	313	-----	312
Db	4790	QRNKAKEVNGKSTKNTITSOLQDYNQLNIAQALRNSVNDVNVNKANSNYINEDNGPKE	4849
Qy	313	-----	312
Db	4850	AYNQAVTHAQTILINAQSNPEMSRDVVNQTKQAVNTAHQNLHGQKLEQAQSSANTEIGNL	4909
Qy	313	-----DSP--MKSIMATQ	324
Db	4910	PNLTNTQAKEKELVNSKQTRTEVQEQALNOAKSLDSSMGTLKSLVAKQPTVQKTSVYINE	4969
Qy	325	-----	324
Db	4970	DQPEQSAYNDSITWGTIINKTADPVLDKTLVDNAISNISTKENALHGEQKLTAKTEAI	5029
Qy	325	-----	324
Db	5030	NALNTLADLNTPOKEAIKTAINTAHTRTDVTASQSKANQINSAMHTLRQNISDNESVTNE	5089
Qy	325	-----	324
Db	5090	SNYINABPEKHAFTEALNNAKEIVNEQATLDANSINQAKAILTTONALDGEQLRRA	5149
Qy	325	-----	324
Db	5150	KENADQEIINTLNLQTDQQRNSEKGLINSQOTRTEVASQLAKAKELNKVMEQLNHLINGKN	5209
Qy	325	-----	324
Db	5210	QMINSSKEFINEDANQQOQAYSNAIASAEALKVKSQNPDLKVTTIEQAINNINSAINNLGE	5269
Qy	325	-----	324
Db	5270	AKLTKAKEDAVASINNLSGLUTNEQTKENQAVNGAQTRDQVANKLRDAEALDQSMQTLRD	5329
Qy	328	LA---TWYTYTLRY-----RQNP-----	342
Db	5330	LVNNQNAIHSSTNYFIEDSTQKNYDINAIDNGSTYITGQNPENLKNSTTDQTSIRINTAK	5389

QY 343 ----- 342
Db 5390 NDHGVKLDKGTANQEIQLGYLNDPKSGEESLVNGSNTRSEVEEHLNEAKSLNNA 5449
QY 343 ----- 342
Db 5450 MKQLRDKVAEKTNVKSSDYINDSTEHQRYDQALQAEANI INEIGNPTLNKSEIBOKLQ 5509
QY 343 ----- 342
Db 5510 QLTDAQNALQSHLLEAKNAITGINKLTALNDAQROKAIENVQAQOTIPAVNQQLTLD 5569
QY 343 ----- 342
Db 5570 REINTAQLRDKVGQONNVHQSNYFNEDEQPKHNDVNSVQAGTTIIDKLODPIMNKNE 5629
QY 343 ----- 342
Db 5630 IEQAINQINTTQALSGENKLHTDQESTNRQIEGLSSLNTAQINAEKOLVNAQAKTRDVA 5689
QY 343 ----- 342
Db 5690 QKLAAAKEINSAMSLNLDGIQNKEDIKRSSAYINADPTKVYTDQALQNAENINATPNV 5749
QY 343 ----- 342
Db 5750 ELNKATIEQALSRVQAQQLDGVQOLANAKQATQTVNGLNSLNDGQKRELNLINSAN 5809
QY 343 ----- 342
Db 5810 TRTKVQBELNKATELNAHEALRNSVQNVQVQKSSNYNVEDQPEQHNYDNVNEAQATI 5869
QY 343 ----- 342
Db 5870 NNNAPVLDKLAIERLTQTVNTTKDALHGAQKLTQDQAAETGIRGLTSLNEPQKAEVA 5929
QY 343 ----- 342
Db 5930 KVTAAATTRDEVRNIROEATTLDTAMLGLRSIKDKNDTKNSKYINEDHDQOQAYDNVN 5989
QY 343 ----- 342
Db 5990 NAOQVIDEQATLSSTINOLANAVQAQSNLHGDTKLQHDKDSAKQTTIAQQLNLSAQK 6049
QY 343 ----- 348
FcPSR-----
Db 6050 HMEDSLIDNESTRTQVQHDLTEAQALDGLMGALKESI KDYTNIVSNGNYINAEPSKKQAY 6109
QY 349 ----- 360
NRTAVSEFMKVT-----
Db 6110 DAAVQNAQNIINGTNOPTINKGNVTATQTVKNTKDALDGHRLBEAKNNANQTI RNLSN 6169
QY 361 ----- 360
Db 6170 LNNAQDAEKVLVNSASTLEQVOQNLQTAQQLDNAMGELRQSI AKKDQVKADSKYLNEDP 6229
QY 361 ----- 360
Db 6230 QIKQNYDDAVQVETIINETONPELLKANIDQATQSVQNAEQALHGAEBKLNQDKQTSSTE 6289
QY 361 ----- 360
Db 6290 LDGLTDLTAQREKLREQINTSNRDDIKQKIEQAKALNDAMKKEQVAQKDGVHANS 6349
QY 361 ----- 360
Db 6350 YTNEDSAQDAYNNALKQAEIDI INNSSNPNLNAQDITNALNNIKQAQDLHGAQLQDDK 6409
QY 361 ----- 360
Db 6410 NTTNQAIGNHLNQPKDALIQAINGATSROQVAEKLEAEALDEAMKQLEQVQVNDQDQ 6469
QY 361 ----- 360

Db 6470 ISNSSPFINESDKQKYNDKIQAAKEIINQTSNPTLDKQKIADTLQNKDANNLHGDQ 6529
QY 361 ----- 360
Db 6530 KLAQSKQDANNQLNLHLDLDTBQONHFKPLINNADTRDEVNKOLEIAKQOLNGDMSTLHKV 6589
QY 361 ----- 360
Db 6590 INDKQIQHLSNVINADNKKQYDNAIKEAEDLIHHPPTLDHKKALQDLNLKIDQAHNE 6649
QY 361 ----- 363
HVL-----
Db 6650 LNGESFKQALDNALNDIDSLNSLVNPPQRTVKDNINHVTTLES LAQELQAKELNDAMK 6709
QY 364 ----- 371
IRNETPYT-----
Db 6710 AMRDSIMNQEQIRKNSNYTNEDLAQNAYNHVAVDKINHIIGEDNATMDPOI IKAQTODIN 6769
QY 372 ----- 371
Db 6770 TAINGLNGDQKLQDAKTDKQITNFTGLTEPQKQALENI INQOTSRAVNAKOLSHAKFL 6829
QY 372 ----- 371
Db 6830 NGKMBELKVAVAKASLVQRNSNYINEDVSEKEAYEQAIKAGOEI INSENNPTISSTDINR 6889
QY 372 ----- 374
IYG-----
Db 6890 TIOEINDAEONLHGDGNKLRQAQOEIAKNEIQNLDSLNSAQITKLIQDGRITTKPAVTQKL 6949
QY 375 ----- 377
TLD-----
Db 6950 EEAKAINQAWQQLKQSIADKDATLNSSNYLNEDSEKKLAYDNVSAEQLEQLINQNDPTMD 7009
QY 378 MSSL----- 381
Db 7010 ISNIOAITOKVIAKDSLHGANKLAQNAQADSNLIIINQSTNLNDKQKQALNDLINHAQTKQ 7069
QY 382 ----- 391
YNETMFVEN-----
Db 7070 QVAEIIAQANKLNEMGTGLTIVEQSNVHQOSKYINEDPOQNIYNDSIQKGREILNGT 7129
QY 392 ----- 391
Db 7130 TDDVLNNKIIADAIQNIHLTKNDLHGDQKLQKAQQDATNELNLYLTNLNSQRQSEHDEIN 7189
QY 392 ----- 391
Db 7190 SAPSRTEVSNDLHAKALNEAMROLENEVALENSVKLSDFINEDBAQNEYSNALQKAK 7249
QY 392 ----- 391
Db 7250 DIINGVPSSTLDKATIEDALLELQNARESLHGEBQKLEAKNOQAVAEIDNLQALNPGQVLA 7309
QY 392 ----- 391
Db 7310 EKTLYNAQSTKEVQEQALQAKELNEAMKALKTEINKKEQIKADSRVYNADSGLOANYNS 7369
QY 392 ----- 400
KTA-----SDSKNT-----
Db 7370 ALNYGSIITATTQPPPELNKDVINRATQTIKTAENNLNGSKLAEAKSGNQSTIEHLQGLT 7429
QY 401 ----- 400
Db 7430 QSQKDKQHDLINQAQTKQOVDDIVNNSKOLDNSMNQLOQI VNNNDNTVKQNSDFINEDSSQ 7489
QY 401 ----- 400
Db 7490 QDAYNHIAQAKDLITAHPTIMDKNQIDQAIENIKQALNDLHGSNKLSEDKKEASFQON 7549
QY 401 ----- 400

Db 7550 LNSLTNGQKDTILNHIAPTRSOVGEKIASAKQLANTMTKALRDSIADNNEILQSSKYFN 7609
QY 401 -----TP----- 402
Db 7610 EDSEQNAYNAVKAKNIINDQPTPWANDEIQSVLNEVKQTKDNLHGDKLANDKTDA 7669
QY 403 ----- 402
Db 7670 QATNALANYNQAQRNLETKVQNSRPEVKVQVLANQLNDAMKGLDALTGNDAIKQ 7729
QY 403 -----TSPM----- 407
Db 7730 TSNYINEDTSQVNFDEYTRGKNIIVAEQTNPNMSPNTINTADKITEAKNDLHGQKLEK 7789
QY 408 ----- 407
Db 7790 QAQQQSINTNQMTGLNQAQKEQLNQEIQOTQTRSEVHVQVINKAQAALNDSMTLRQSITD 7849
QY 408 ----- 407
Db 7850 EHEVKQTSNYINETVGNQTAYNNAVDVKQIINQTSNPTMNPLEVERATSNVKISKDALH 7909
QY 408 ----- 407
Db 7910 GERELNDKNKSTFAVNLHDLNLAQAQKEALTHEIEQATIVSQVNNIYNKAKALNNDMKKL 7969
QY 408 ----- 407
Db 7970 KDIVAQODNVRQSNYINEDSTPQNMVNDTINHAQSIIDQVANPTMSHDEIENAINNIKH 8029
QY 408 ----- 407
Db 8030 AINALDGEHKLQAQKENANALLINSLDNLAPORDAINRLVNEAQTRKVAEQLSQAALN 8089
QY 408 ----- 407
Db 8090 DAMKHLRNSIQNSSVRQESKYINASDAKKEQVNHAVREVENIINEQHPTLDKEIYKQLT 8149
QY 408 ----- 407
Db 8150 DGVQANNDLNGVELLDADQNAHQSIPTLMLHLNQAQNALNEKINNVAITRTEVAALIGQ 8209
QY 408 ----- 407
Db 8210 AKLLDHAMENLEBSIKDKEQVQSSNYINEDSDVQETYNDAVDHVTEILNQTVNPTLSIE 8269
QY 408 ----- 407
Db 8270 DIEHAINEVNAQKQLRGKQKLYQTTIDLADKELSKLDDLTSQQSSSISNQIYTAKTRETV 8329
QY 408 ----- 407
Db 8330 QAQIEKAKSLNHAMKALNKVYNADKVLDSRFINEDQPEKKAYQQAINHVDISIHRQTN 8389
QY 408 ----- 407
Db 8390 PEMDPTVINSITHELETAQNHLHGDKLAHQQAANVINGLIHLNVAQREVMTNTNA 8449
QY 408 ----- 407
Db 8450 TTREKVAKNLDNAQALDKAMETLQQVVAHQKNILNDSKYLNEDESKYQQQYDRVIADAEQL 8509
QY 408 ----- 407
Db 8510 LNQTTNPTLEPYKVDIVKQNVLANEKILFGAEKLSYDKSNANDEIKHMYLNNAQKQSIK 8569
QY 408 ----- 407
Db 8570 DMISHAALRTEVKQLLQQAAILDEAMKSLDKTQVWITDTPLNTEASEDKKEKVDQTV 8629
QY 408 ----- 407
Db 8630 SHAQAIIIDKINGSNVSLDQVRQALEQTOASENLGDGQDRVEEAKVHANQOTIDQLTHLSL 8689

QY 408 ----- 407
Db 8690 QOQTAKESYKNAIKLEESITATSVNNAQALNKVMGKLEQFINHADSVENSQNYROADDKII 8749
QY 408 -----GFOR----- 411
Db 8750 AYDEALEHGDIOKTNATONETKQALQOLIYAETSINGFERLHNHARPRALEYIKSLEKIN 8809
QY 412 ----- 411
Db 8810 NAKSALEDKVTQSHDILLELEHIVNEGTVNLDMGELANAIVNNYAPTASINYINADNL 8869
QY 412 -----TFIDPLW----- 421
Db 8870 RKDNFTQAINNARDALNKTKQGNLDFNAIDTTFKDDIFKTKDALNGIERLTAASKAKAEKLI 8929
QY 422 DSIILFL-----DEIRNF-----SLR-----SPT 439
Db 8930 DSLKFINKAQFTHANDEIMNTNSIAQLSRIVNOAFDLNDAMKSLRDELNNQAPFVQASSN 8989
QY 440 YVN----- 442
Db 8990 YINSEDELKQFDPHALSNARKVLAKENGKVLDEKQIOGLKQVIEDTKDALNGIQRLSKAK 9049
QY 443 ----- 442
Db 9050 AKAIQVVSLSYINDAQRHIAENNIHNSDDLSSLANLTSKASDLONAMKDLRDTIESNST 9109
QY 443 ----- 442
Db 9110 SVPNVNYINADKNLQIEPDEALQOASATSSKTSNPATIEEVLGUSQAIYDTGNALNGE 9169
QY 443 ----- 442
Db 9170 QRLATEKSKDLKLIKGLKDLNKAQLEDVTNKVSANTLTLSQLTQSTLELNDKMKLLRD 9229
QY 443 -----LTPP----- 446
Db 9230 KLKTLVNPVKASLYRNADYNLKRQFNKALKEAGVLNKNSGTNNVINDIOHLLTQIDNA 9289
QY 447 -----EHR----- 450
Db 9290 KDQLNGERLKEHQKSEVFIKELDILNNAOKAAIINOIRASKDIIKINOIVDNAIELN 9349
QY 451 -----AVNL----- 454
Db 9350 DAMQGLKEHVAQITATTKDNIEVYLNADHKLQYDYAINLANNVLDKENGTKNDANIIG 9409
QY 455 -----STNS----- 459
Db 9410 MIQNMDARALLNGIERLKDAQTKAHNDIKDTLKRQLEBIEHANATSNKSAQAKQMVNEE 9469
QY 460 ----- 459
Db 9470 ARKALSNTINDATSNDLVNOAKDEGOSAIIEHIADELPHAKLDANQMIQKVEDINHLISQ 9529
QY 460 ----- 459
Db 9530 NPNLSNEEKNKLI SQINKLVNGIKNEIQQAINKQOQIENATTKLDDEVIEETTKLIIAKAEA 9589
QY 460 ----- 459
Db 9590 KQMIKELSQKRDAINNNNTDLTPSKAHALADIDKTEKDALQHIENSNSIDINNNEHA 9649
QY 460 -----LW----- 461
Db 9650 FNTLAHIIITDQDQPLVFEVPELSLQNALVTSVVVVRDETISLESIKKQMTLTDLKV 9709
QY 462 ----- 461
Db 9710 NIVSLPNTDKVADHLTAKVKVILADGSYTVNVVPKVVKELOIAKDAIKTIDVLVKQK 9769

```
Qy 462 ----- 461
Db 9770 IKDIDNNELTSTOEDAKAERLKQKQIDKVNHSKSIKDIETVKRTDPEIDQDPKR 9829
Qy 462 ----- 461
Db 9830 PTLNKAKKDIIITDNTIOINGPKIEITIKGLTSNEKTQFDKQLTALQKEFLEKVEHAHL 9889
Qy 462 ----- 461
Db 9890 VELNQLQOEFNNRYKHILNQAHLLGEXKHIAEHKLGIVVNNKTOQILNNQSASYFIKWAL 9949
Qy 462 ----- 461
Db 9950 DRIKIQIETWNSIRGAHTVQDVHKKALLQIEQILKVNVSIIINQSFNDSLHNFYLHSKF 10009
Qy 462 ----- 461
Db 10010 DARIREKDVANHIVQTFEVLKGTGVEPKINKETQQPKLHKNDNDSLFKHLVDNFGK 10069
Qy 462 -----WWL 464
Db 10070 TVGVITLTGLSSFWL 10085

RESULT 4
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

Query Match 19.7%; Score 483; DB 3; Length 11877;
Best Local Similarity 2.0%; Pred. No. 0.0013;
Matches 229; Conservative 77; Mismatches 149; Indels 11020; Gaps 74;

Qy 2 GRKEMVRDVPKM----- 14
Db 217 GLAEILRRHIPKRTYRSLADRHGIPDGGLPKDLTLTQWIALFOASQPSYAPGAPGTRMPG 276
Qy 15 ----- 14
Db 277 QGGAGGRDYDSETSRAAVGSSRRYGTTRGGEPCAPRAQVRQTKGRQARGSSYGRRTGR 336
Qy 15 ----- 14
Db 337 MSSAGITRTGARTPTVTRGNAAWDTGEVRVRRLPPAGPDHAEHSFSAFTGDRVRAELIR 396
Qy 15 -----FVLIS----- 19
Db 397 GEMSTVSKSESEBFVSNDAGSAHGTAEFPVAVVGISCRVPGARDPREFWELLAAGQAV 456
Qy 20 ----- 19
Db 457 TDVPADRWNAGDFYDPRSAGRNSRWGGFTBEDVDRFDAAPFGISPREAAEMDPQORLA 516
Qy 20 ----- 19
Db 517 LEIGWEALERAGIDPSSLTGTGTGVPAGAIWDDYATLKHROGGAITPHTVTGLHRGIITA 576
```

```
Qy 20 -----ISFLV----- 25
Db 577 NRLSYTLGLRGSVMVDSQSSSLVAHVHLACESLRGESELALAGGVSLNVLNLPDSIGAS 636
Qy 26 -----SFINCKVMSKAL----- 37
Db 637 KFGGLSPDGRAVTFDARANGYVRGEGGFVWLKRLSRAVADGDPVLAVIRGSVAVNGGAA 696
Qy 38 ----- 37
Db 697 QGNTTPDAQOEAVALREAHERAGTAPADVRYVELHGTGTPVGDPIEAAALGAALGTGRPA 756
Qy 38 -----YNNP----- 41
Db 757 GQPLLGVSVKTNIGHLEGAAGTAGLIKAVLAVRGRALPASLNYETENPAIPFEELNLRVN 816
Qy 42 ----- 41
Db 817 TEYLPWEPEHGDQRMVVGVSFQMGGTNAHVLEAPGCGRGASVVESTVGGSAVGGVV 876
Qy 42 ----- 41
Db 877 PWVSAKSAALDAQIERLAASFASRDTDGDAGAVDAGAVARVLAGRAQFEHR 936
Qy 42 -----WRGL----- 45
Db 937 AVVVGSGPDDLAALAAPEGLVRGVASGVGRVAFVPPGQGTQWAGMGAELLDSSAVFAA 996
Qy 46 ----- 45
Db 997 MAECAALSPYVDWSLEAVVRQAPGAPTLERVDVVPVTFVAVVSLARVMQHHGVTPOAV 1056
Qy 46 -----VLSKIGY 53
Db 1057 VGHSGEIAAAVAGALSDDAARVVTLRSKSIAAHLAKGKMWLSLASEDAVLERLAGF 1116
Qy 54 -----KLDQKLEIL-----RQLE----- 67
Db 1117 DGLSVAAVNGPTATVVSGDPVQIEELARACEADGVRARVIPVDYASHSQVEIESELAE 1176
Qy 68 -----TTISTK-----Y 74
Db 1177 VLAGLSPOAPRVPPFFSTLEGAMITBPVLDGGYVYRNLRHVRGFAPAVETLATDEGTFHV 1236
Qy 75 NVSKOPVKNLTMN----- 87
Db 1237 EVSAHPV--LTWALPGTVTGLATLRDNGGQDRLVASLAEAWANGLAVDWSPLPSATGH 1294
Qy 88 ----- 87
Db 1295 HSDLPTYAFQTERHWLGEIEALAPAGEPAVQPAVLRTAEABEPAELDRDQLRVLDKVRA 1354
Qy 88 ----- 87
Db 1355 QTAQVLGYATGQIEVDRTFREAGCTSLTGLDRLNRINAAFGVRMAPSMIFDFTPEALA 1414
Qy 88 ----- 87
Db 1415 EQLLLVHGEAANAAPAGAPAPVAAAGAVEPVAIVGMACRLPGGVASPEDLWRLVAGGG 1474
Qy 88 -----TEFPQ----- 92
Db 1475 DAISEFPQDRGWDVGLYHPDPEHPGTSYVRQGGFIENVAGFDAAPFGISPREALAMPQ 1534
Qy 93 -----YYILAG----- 98
Db 1535 QRLLETSWEAVEDAGIDPTSLRGRQGVFTGAMTHEYGPSLRDGGEGDGLYLLTGNTAS 1594
Qy 99 ----- 98
Db 1595 VMSGRVSYTLGLEGPALTIVDTACSSSLVALHLVAQLRKGEVDMALAGVAVMPTGMFV 1654
```

QY 99 ----- 98
Db 1655 BFSRQRLAGDGRSKAPASADGTSWSEGVLLVERLSDARNRNGHVLAVRGSAHQD 1714
QY 99 -----PI----- 100
Db 1715 GASNGLTAPNGSQORVIRPALADARLTTSDDVDVEAHGTGTLGDPFIEAQALITYGOG 1774
QY 101 ----- 100
Db 1775 RDDEQPLRLGSLKSNIGHTQAAAGVGVIMQVQAMRHGLLPKTLHVDEPSDQIDWSAGAV 1834
QY 101 ----- 100
Db 1835 ELLTEAVDWPEKODGGLRAAVSSFGISGTNAHVLEAPVVPVVEGASVVEPSVGSAGVG 1894
QY 101 ----- 100
Db 1895 GVTTPWVSAKSAALDAQIERLAASFASRORTDDADAGAVDAGAVAHVLADGRAQFEHRAV 1954
QY 101 ----- 100
Db 1955 ALGAGADDLVQALADPDGLIRGTASGVRVAFVPPGQGTQWAGMGAELLDSSAVFAAAMA 2014
QY 101 -----QYYSIT----- 106
Db 2015 ECEAALSPYVDMNSLEAVVRQAPGAPTLEVVVQVPTFVPMVSLARVWQHGVTPQAVVG 2074
QY 107 ----- 106
Db 2075 HSQGEIAAYVAGALPLDDAARVVTLRSKSIAAHLAKGKMSLALNEDAVLERLSDPDG 2134
QY 107 ----- 106
Db 2135 LSVAAVNGPTATVSGDPVQIEBELAQACKADGPRARIIPVDYASHRSQVEIIESELAQVL 2194
QY 107 ----- 106
Db 2195 AGLSPOAPRPPFSTLEGTWITEPVLDTGYWYRNLRRHVGFAPAIETLAVDEGTFHFEV 2254
QY 107 ----- 106
Db 2255 SAHPVLTMTLPETVTLGLTLRREGQGERLVTLSAEAWNGLPVAWTSLLPATASRPLP 2314
QY 107 ----- 106
Db 2315 TVAFQERYWLENTPAALATGDDWRYRIDWKRLPAEAGSERTGLSGRWLAVTPEDHSAQA 2374
QY 107 ----- 106
Db 2375 AAVLTALVDAGAKVEVLTAGADDDREALAARLTALTGDTGTVVSLDGLVPQVAWQA 2434
QY 107 ----- 106
Db 2435 LGDAGIKAPLNSVTQASVSGRLDTPADPDRAMLWGLGRVVALEHPERWAGLVDLPQAPD 2494
QY 107 ----- 106
Db 2495 AAALAHVLTALSGATGEDQIAIRTTGLHARRLARAPLHGRRPTRDMQPHGTVLITGGTGA 2554
QY 107 ----- 106
Db 2555 LGSHAARMMAHGAHELLLVSRSGEQAPGATQLTABLTASGARVTIAACDVADPHAMRTL 2614
QY 107 ----- 106
Db 2615 LQIPAEETPLTAVVHTAGALDDGIVDTLTABQVRRRAHRAKAVGASVLDELTRDLDLDAFV 2674
QY 107 -----YL----- 108
Db 2675 LFSSVSTLIGIPCGNYAPHNAVLDALAAARRATGRSAVSVANGPHDGGGMAAGDQVAER 2734
QY 109 -----WDFY----- 113

Db 2735 LRNHGVPQMDPELALAALESALGRDETAITVADIDMDRFTLAYSGRPOPLVEBELPEVRR 2794
QY 114 ----- 113
Db 2795 IIDARDSATSGOGSSAQGANPLAERLAAAAPCERTEILLGLVRAQAAAALVLRMRSPEDVA 2854
QY 114 -----STQLRKPKYVY----- 125
Db 2855 ADRAFMDIGFDSLAGVELNRNLTRATGLQPATLVFDPHTPLALVSLLRSEFLGDEBTAD 2914
QY 126 ----- 125
Db 2915 ARRSAAALPATVAGAGAGAGAGTADDPIAIVAMSCRYPGDIRSPEDLWRMLSEGGSGITP 2974
QY 126 ----- 125
Db 2975 PFTDRGMDLGLYDADPDALGRAYVREGGFLHDAAEFDFGVSPREALAMPQORMLL 3034
QY 126 ----- 125
Db 3035 TTSWEAFERAGIEPASLRSGSTGVFIGLSYQDYAARVPNAPRGVEGYLLTGSTPSVASGR 3094
QY 126 ----- 125
Db 3095 IAYTFGLEPATVTDACSSLTALHLAVLRALRSRECTMALAGGVAMMATPHMFVFSRQ 3154
QY 126 ----- 125
Db 3155 RALAPDGRSKAFSADADGFGAAEGVCLLVERLSDARRNGHPVLAVVGTAVNQDGASNG 3214
QY 126 ----- 125
Db 3215 LTAPNGSQORVIRQALADARLAPGDIDAVETHGTGTSIGDPTEAQGLQATYKERPAAER 3274
QY 126 -----SQYNT----- 131
Db 3275 PLAIGSVKSNIGHTQAAAAGAGIIRKWLAMRHGTLPKTLHADEPSPHVDWANSGLALVTE 3334
QY 132 ----- 131
Db 3335 PIDWPAGTGPRAAVSSFGISGTNAHVLEQAPDAAGEVLGADEVPEVSETVAMAGTAGT 3394
QY 132 ----- 131
Db 3395 SEVAEGSEASEAPAPGSGREASLPCHLPWLAKDBEQSLRGQAAALHAWLSEPAADLSDA 3454
QY 132 ----- 131
Db 3455 DGPARLRDVGYTTLATSRTAFAHRAAVTAADRGFLDGLATLAQGGTSAHVHLDLTARDGTT 3514
QY 132 ----- 131
Db 3515 AFLPTQGSQORPGACRELYDRHPVAFARALDEICAHLDGHLELPLLDVMPFAEGSAEAAALL 3574
QY 132 ----- 131
Db 3575 DETRYTQCALFALEVALFRLVESWGMPPAALLGHSVGEIAAAHAGVAGVFLSADAARLVAAR 3634
QY 132 ----- 131
Db 3635 GRLMQELPAGGAMLAHQAAADEIRVWLETFEYRAGRLDVAAVNGPEAAVLSGDADAAREA 3694
QY 132 -----AKTITPRPP----- 140
Db 3695 EAYNSGLGRTRTALRVSHAFSAHMDGMDGDFRAVLTEVFEFRPRLTTVVSVNTGLAAGPD 3754
QY 141 ----- 140
Db 3755 DLCDPEYVVRHVGTVRFLDGVRLDLGVRTCLELGPDLVTAMAADGLADTADTADSAAG 3814
QY 141 ----- 140

Db 3815 SPVGPAGSPADSAAGALRPRPLLVALLRRKRSETETVADALGRAHAHGTGPDWHAWFAG 3874
Qy 141 ----- 140
Db 3875 SGAHRVLDPTYSPFRDRYLDAPAADTAVDTAGLGTADHPLLGAVVSLPDRDGLLLTG 3934
Qy 141 ----- 140
Db 3935 RLSLRTHPLADHAVLGSVLLPGAAVVELAAHAESAGLRDVRELTLLEPLVLPHEGGVE 3994
Qy 141 -----PCGR----- 144
Db 3995 LRVTVGAPGPGESAGDGARPVSLHSLRLADAPAGTAMSWCHATGLLATDRPELPVAPDR 4054
Qy 145 ----- 144
Db 4055 AAMWPPQAGEVPLDGLYERLDGNGLAFGPLFQGLNAVRYEGEVFADIALPATTNATAP 4114
Qy 145 ----- 144
Db 4115 ATANGGSAAPYGIHPALLDASLHAIAVGLVDEPELVRVPFWHSGVTVAAGAAAAAR 4174
Qy 145 ----- 144
Db 4175 VRLASAGTAVSLSLTDGGRPLVSVVERLTLRPVTADQAAASRVGGLMHRVAMRVPYALAS 4234
Qy 145 -----VPSMTCLSE----- 153
Db 4235 SGBQDPHATSYGPTAVLGKDELKVAALASAGVEVGLYPLDALSLQDVAAGAPRPTVLA 4294
Qy 154 ----- 153
Db 4295 PLPAGPADGAGVGRTVARTLELQAWLADEHLAQLLTVRGAVRDPGSGADGGE 4354
Qy 154 ----- 153
Db 4355 DLSHAAWGLVRTAQTENPGRFGLLDLADDASSYRTLPSVLSDAGLRDEPQLALHDGTTIR 4414
Qy 154 -----MLNYSKR----- 160
Db 4415 LARLASVRPETGTAAAPALAPEGTVLLTGGTGGGLGVARHVWGEWVRLLLVSRRGTD 4474
Qy 161 ----- 160
Db 4475 PGADLVHELEALGADVSAACDVADREALTAVLDAIPAEPHPLTAVVHTAGVLSGDTLPS 4534
Qy 161 ----- 160
Db 4535 MTTEDVEHVLRPKVDAAPFLDLDELSTPAYDLAAPVMPSSAAAVFGAGQGAYAAANATLD 4594
Qy 161 ----- 160
Db 4595 ALAWRRRAAGLPALSLGWLAETSMTGELGQDLRRMSRAGIGGISDAEGIALLDLAL 4654
Qy 161 -----ND----- 162
Db 4655 RDRHPVLLPLRLDAAGLRDAAGNDPAGIPALFRDVGARTVRRPSSAASSTAGTAGT 4714
Qy 163 ----- 162
Db 4715 PGTDGAETAATVLADRAATVDGPARQRLILLEFVVVGEVAEVLGHARGHRIIDAERGFLDL 4774
Qy 163 ----- 162
Db 4775 GFDSLTAVELNRNLNSAGGLPATLVFDHPSPALASHLDLAEPRGASDQDAGNRNGN 4834
Qy 163 -----TGQGGCN 170
Db 4835 ENGTTASRTAETDALLAQLTRLEGALVLTGLSDAPGSEVLEHLRSLRSMVTGTGTGT 4894
Qy 171 FT----- 172
Db 4895 ASGAPDGAGSGBDRPWAAGDGAGGSGEDGAGVPDFPMNASAEELFGLLDQDPSTDMSTVN 4954

Qy 173 ----- 172
Db 4955 EEKYLDIRATADLHEARGRLRELEAKAGEPVAIVGMACRLPGGVASPEDLURLVAGGE 5014
Qy 173 -----TFNPMFFNV-PR----- 183
Db 5015 DAISEFPQDRGMDVEGLYDPNPEATCKSYAREAGFLYEAGEFDADFFGISPREALAMDPC 5074
Qy 184 -----WN-----TKL-----YVGP----- 192
Db 5075 QRLLEASWEAFEHAGIPAAATARGTSVGVTGMVMDHYATRLTDVPEGIEGLTNGSGS 5134
Qy 193 ----- 192
Db 5135 VASGRVAYTLGLEGPVTVDTACSSSLVALHLAVQALRKGEVDMALAGGVTVMSTPSTFV 5194
Qy 193 -----TKVNV 198
Db 5195 EFSRQRLAPDGRSKSFSTADGTSWSEGVGLLVERLSDARRKGHRILAVVVGTAVNQD 5254
Qy 199 SQTIFYPLGLTA----- 209
Db 5255 GAS---SGLTAPNGPFSQORVIRRALADARLTTSDDVDVBEAHGTGTRLGPPIEAQVIATY 5311
Qy 210 ----- 209
Db 5312 GQGRDGEQPLRLGSLKSNIGHTQAAAGVSGVIMVQAMRHGVLPKTLHVKEKPTDQDWSA 5371
Qy 210 ----- 209
Db 5372 GAVELLTEAMDWPDKDGLRRAAVSSFGVSGTNAHVLEEAPAAEETPASEATPAVEPS 5431
Qy 210 ----- 209
Db 5432 VGAGLVPMVLSAKTPAALDDAQIGRLAAAFASQGRDADDPGARVARVLAGGRABFEHRAVVL 5491
Qy 210 ----- 209
Db 5492 GTGQDDFAQALTAPEGLIRGTPSDVGRVAFVFPQGTQWAGMGAEILLDYSKEFAAAMAEBC 5551
Qy 210 ---LLRYAQ----- 216
Db 5552 ESALSRYTWSLEAVVROAPGAPTILERVVDVQVPTFAVMVSLAKVWQHGVTPQAVVGH 5611
Qy 217 ----- 216
Db 5612 QGEIAAAVAGALTLDADAARVVTLRSKSIAAHLAKGGMI SLASEEATRQRIENLHGLS 5671
Qy 217 ----- 216
Db 5672 IAAVNGPTATVSGDPTQIQELAQACEADGVRARIIPVDYASHSAHVETIESELAELVLAG 5731
Qy 217 -----RNCNH----- 221
Db 5732 LSPRTPEVPFSTLEGAWITEPVLDTGYWRNLRRHVRGFAPAVETLATDEGTFHEIVSA 5791
Qy 222 ----- 221
Db 5792 HPVLTMTLPTETVTLGTLTBRREQGQERLVTSLAEAWTNGLTIDWAPVLPTATGHHPELPT 5851
Qy 222 ----- 221
Db 5852 YAFQRRHYLHDSPAVQGSVQDSWRYRIDWKRLAVADASERAGLSRVLVVPEDRSAEA 5911
Qy 222 ----- 221
Db 5912 APVLAALSGAGADPVQLDVSPLGDRQRLAATLGEALAAAGGAVDGVLSLLAWDESAHPGH 5971
Qy 222 -----SFLVNA----- 228
Db 5972 PAPFTRGTGATLTLVQALEADAGVAAPLWCVTTHGAVSVGRADHVTSPAQAMVWGMGRVAAL 6031

QY 229 -----MSRNLFRVPKYNGTK 244
DB 6032 EHPERWGLIDLPDADRALDRMTTVLAGGTGEDQVAVRASGLLARLRVLRASLPAGHTA 6091
QY 245 ----- 244
DB 6092 SPWQADGTVLVTCABEPAAAEARRLARDGAGHLLHTTPSGEAGBTSGAEDSGLA 6151
QY 245 ----- 244
DB 6152 GLVAELADLGATATVTCDLTDAEAAARLLAGVSDAHLPLSAVLHLPTVDSEPLAATDAD 6211
QY 245 ----- 244
DB 6212 ALARVVAKATAHLHLDRLLREAAAGRPVVLVLFSSVAIIMGAGQAGYAAGTAFLLDA 6271
QY 245 ----- 244
DB 6272 LACQHRADGPTVTSVAWSPEWGSRVTEGATGERLRLGLRPLAPATALTALDHALGHGDT 6331
QY 245 ----- 244
DB 6332 AVTIADVMSFPAGFTTARPQTLLADLPEARALDEQQSTTAADTVLSRELALGTAE 6391
QY 245 ----- 244
DB 6392 QORRMQELVREHLAVVLNHPSPPEAVDTGRAFRDLGFDLSLTAVELNRNLKNATGLALPATL 6451
QY 245 ----- 244
DB 6452 VFDYPTPTLAEFLAELIGEQAAGEQLPVDGGVDPEVAIVGMACRLPGGVASPEDLM 6511
QY 245 ----- 244
DB 6512 RLVAGGEDAISGPPQDRGWDVEGLYDPPDASRTYCRAGGLDEAGEFDADFFGISPRE 6571
QY 245 -----LKNTRKJK----- 253
DB 6572 ALAMDPORLLETSWEAVEDAGIDPTSLQGGQGVGFAGTNGPHYBPLLRNTAEIDLEGYV 6631
QY 254 ----- 253
DB 6632 GTGNAASIMSGRVSYTLGLEGPATVDTACSSLSVALHLAVQALRKGEGLALAGGVTM 6691
QY 254 ----- 253
DB 6692 STPTTFEFSRQGLAEDGSKAFASADGFGPAEGVGMLLVERLSDARRNGHRVLAVR 6751
QY 254 ----- 253
DB 6752 GSAVNDGASNGLTAPNGPSQQRVIRRALADARLTADVDVVEAHGTGRLGDPIEAQAL 6811
QY 254 ----- 253
DB 6812 IATYQGRDTEQPLRLGSLKSNIGHTQAAAGVSGIIMVQAMRHGVLPKTLHVDPSDQI 6871
QY 254 -----RKQ-----APVKE----- 261
DB 6872 DMSAGTVELLTEAMDPKQEGGLRRANVSSFGISGTNAHIVLEAPVDEDDAPADEPSVG 6931
QY 262 -----QFEKKA----- 267
DB 6932 GVVPLVLSAKTPAALDAQIGRLAASFASQGRTTDAADPGAVARVLAGGQAQFEHRAVALGTG 6991
QY 268 ----- 267
DB 6992 QDDLAALAAPEGLVRGVASGVRAFPVPGQGTOWAGMGAELLDVSKFAAAMACEAA 7051
QY 268 ----- 267
DB 7052 LAPYDWSLEAVVRQAPGAPTLEERVDPVPTFAMVMSLAKVWQHGHGVTFOAVVGHSGQE 7111
QY 268 ----- 267

DB 7112 IAAAVAGALSDDAARVVVTLRSKSTGAHLAQGGMLSLALSEAAVVVERLAGFDGLSVAA 7171
QY 268 -----KKT 270
DB 7172 VNGPTATVSGDPTQIQELAQACEADGVRAIIPVDYASHSAHVETIESELADVLAGLSP 7231
QY 271 QSTTTPYFS----- 279
DB 7232 QTPQVFPFSTLEGAMITEPALDGGWYRNLRHRVGPAPAVETLATDEGTFHFEVESAHPV 7291
QY 280 -----YTS-----HALNVTNV-----TYSIT 297
DB 7292 LTMALPETVTGLURRDNGGQHLRTSLAEAWANGLTVDWASLLPTTTTHDPLDYAFQ 7351
QY 298 T----- 298
DB 7352 TERYWQPDLSAAGDITSAGLGAEBPLLGAVALADSDGCLLTGSLSLRTHFWLADHAV 7411
QY 299 ----- 298
DB 7412 AGTVLLPGTAPVELAFRAGDQVGC DLVEBLTLDAPLVLPERRGAVRVQLSVGASDESGRRT 7471
QY 299 -----AARRVSTSTIA----- 309
DB 7472 FGLYAHPEADAGEAEWTRHATGVLAAADRTAPVADPEAWPPPGAEFPVDVGLYERFAAN 7531
QY 310 ---YRP----- 312
DB 7532 GYGYGLFGVGVWRGDEVFADVALPAEVAGAEGARFGLHPALLDDAAVQAAGAGRGVR 7591
QY 313 -----DSS----- 315
DB 7592 RGHAAAVRLERDLLYAVGATALRVLAPAGPDTVSVSAADSSQGPVFAADSLTVLPVDPA 7651
QY 316 ----- 315
DB 7652 QLAAFSDPTDLALHLEWTAWDGAQAALPGAVVLGGDADGLAAALRAGGTEVLSFPDLTD 7711
QY 316 ----- 315
DB 7712 LIVEAVDRGETPAPATVLVACPAAGPDGPEHVREALHGLMALMQWLADERFTDGLRLVLT 7771
QY 316 ----- 315
DB 7772 RDAVAARSGLRSTGQAAMVGLRSGAQTESPGRFVLLDLAGEARTAGDATAGDGLTTGD 7831
QY 316 ----- 315
DB 7832 ATVGTTSGDAALGSALATALGSPEQLALRDGALLVPRLARAAAPAAADGLAAADGLAAL 7891
QY 316 ----- 315
DB 7892 PLPAAPALWRLBPGTDGSLSLTAAPGDAETLAPBPLGPGQVRIATRATGLNFRDVLIAL 7951
QY 316 ----- 315
DB 7952 GMYPDALMGTEGAGVVTATPGVTHLAPGDRVMGLLSGAYAPVVVADARTVARMPEGWT 8011
QY 316 -----FMKSTMATOLRDLA----- 329
DB 8012 FAQGASVPVVFULTAVYA--LRDLADVKGPERLLVHSAAGVGMAAVQLARHMGVEVHGTA 8069
QY 330 ----- 329
DB 8070 SHGKWDALRALGLDDAHIASSRTLOPFESAFAASGGAGMDVVLNLSAREFVDASLLGLGP 8129
QY 330 ----- 329
DB 8130 GGRFVEMGKTVDYDAERVAADHPGVYRAFDLGEAGPERIGEMLAEVIALFEDGVLRLHP 8189
QY 330 ---TW----- 331
||

Db	8190	VTTWDRRARDAFRHVSQARHTGKVLTMPSGLDPEGTVLLTGTGTGALGGIVARHVVGWE	8249	Qy	368	----	TPYTIYG	----	TLD	----	MSSLYYNE	385				
Qy	332	----	----	Db	9330	QHGGITTPQAVGVHSGQGEIAAAVYAGALTLLDDAAVVVTLRSKSIAAHLAKGKGMISALDE	9389	Qy	386	----	----	385				
Db	8250	GVRELLVSRGTDAPGAGELVHELEALGADVSAACDVADREALTAVLDSIPAHEPLTA	8309	Db	9390	AAVLKELSDPDLGSVAAVNGPTATVVGSDPTQIEELARTCEADGVPARIIPVDYASHSRQ	9449	Qy	386	----	----	385				
Qy	332	----	----	Db	9450	VEIIEKELAEVLAGLAQAPHPVFPFSTLEGTWITEPVLDTYWRNLRRHVRGFAPAVETL	9509	Db	386	----	----	385				
Db	8310	VVHTAGVLSGTLPSMTAEVDEHVRPKVDAAFLDLSTPGVDLAAPFVWFSSAAVFG	8369	Qy	386	----	TMFVE	----	----	----	----	390				
Qy	336	----	----	Db	9510	AVDGFTHFIEVSAHPVLTMTLPETVTGLTLRREGQGERLVTSLAEAWANGLTIDWAPI	9569	Qy	391	----	----	390				
Db	8370	GAGQAGYAAANATLDALAWRRRTAGLPALSLGWLMAETSGMTGSLDTRSLARSGAT	8429	Db	9570	LPATGHHPELTYAFQTERFVLQSSAPTSAAEDMRVYRVEWKPLTASQADLSGRWIVAV	9629	Qy	391	----	----	390				
Qy	336	----	----	Db	9630	GSEPEAELGALKAAAEVDVLEAGADDDREAAARLTALTITGDTGFTGVVSLDDLDPQV	9689	Qy	391	----	----	390				
Db	8430	PMDELTLISLDDAAMRDPALVPIALDVAALRAQQRDGLMLPLLSGLTRGSRVGAPVN	8489	Db	9690	AMVQALGDAGIKAPLWSVTQGAHSVGRDLTPADPDRAMLNGLGRVVALEHPERWAGLVDL	9749	Qy	391	----	----	390				
Qy	336	----	----	Qy	391	----	----	Qy	391	----	----	390				
Db	8490	QRRAAAGGAGEADTLGGRLAAMTPDDRVAHLRDLVRTHVATVLGHGTPSRVDLERAFRD	8549	Db	9750	PAQPDAAALHLVTALSGATGSDQIAIRTTGLHARRLARAPLHGRRTDQMOPHTVLIT	9809	Qy	391	----	NKTASDSNKT	400				
Qy	336	----	----	Qy	391	----	----	Db	9810	GGTGALGSHAARWMAHGAELLLVSRSGEQAPGATQLTAEILTASGARVTIAACDVAPH	9869	Qy	401	----	----	400
Db	8550	TGFDSLTAVELRNLNAATGLRLPATLVFDHPTFGLAGHLDELATAAGGSWAEGTSG	8609	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	342	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8610	DTASATDQTTAALAELELDLEGVLASLAPAAGRPPELAARLALAAALGDDGDATDLDE	8669	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	342	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8670	ASDDDLFSFIDKELGSDFMANNEDKLRYLKRVTAELOQNTLRLEIEGRTHEPVAIVG	8729	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	342	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8730	MACRLPGVASPEDLWQVAGDGAISEFPQDRGWDVEGLYDDPDASGRTYCRSGGFLH	8789	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8790	DAGEFDADFGISPREALAMDPOQLSLTTTAWAIESAGIDPTALKSGSLGVFVGWHGTG	8849	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8850	YTSGOTTAVQSPLEGLHVSAAALGFLSGRIAYVLGTGDPALTVDTCSSSLVALHLAVQ	8909	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8910	ALRKGCDMALAGVTVMFNADLFVQFSRQRLAADGRSKAFATSDGFGPAEGAGVLLV	8969	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	8970	ERLSDARRNGHRILAVVSGAVNQDGSNGLTAPHGPSQQRVIRRALADARLAPGDVDVV	9029	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	9030	EAHGTGRLGDPTEAQAALTIATYQEKSEBQPLRLGALKSNIGHTOAAAGVAGVIRKVOAM	9089	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	345	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	9090	RHCLLPKTLHVDPSQIDWSAGTVELLTEAVDWPKEQDGLRRAAVSFGISGTAHVV	9149	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	363	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	9150	LEEAPAVEDSPAEPAGGVWPVPSAKTPAALDAQIGLAAAYADRTDVPAAARAL	9209	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	363	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	9210	VDSRTAMEHRAVAGDSREALRMLPPEGLVRGTSSDVRVAFVFPQGTQWAGMGAE	9269	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Qy	363	----	----	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390
Db	9270	LDSSPEFRASMAECETALSRYVDWSLEAVVRQEPGAPTILDRVDVVQPVTFVAVWSLAKW	9329	Qy	391	----	----	Qy	401	----	----	Qy	391	----	----	390

QY 434 ----- 433
Db 10410 QYAPDIRVAPEGTGGYVVTGNSSAVASGRISAYSLGLEGPAVTVDTACSSSLVALHLAK 10469
QY 434 ----- 433
Db 10470 GLRNGDCSTALVCGVAVLATPGAFIEFSSQQAADGRKYGFAAADGLANGEGVAVLLL 10529
QY 434 ----- 433
Db 10530 ERLSDARRKGRHVLAVRGSAINQDGSNGLTAPHGSPQHLIRQALADARLTSSDDVV 10589
QY 434 ----- 433
Db 10590 EHGHTGRLGDPTEAQALLATYQGRAPQGLRLGLTKSNIGHTQAASGVAGVVKMVQAL 10649
QY 434 ----- 433
Db 10650 RHGVLPKTLHVDEPTQVDWSAGSVELLTEAVDWPBPGRLRAGVSAGVGGTNAHVVL 10709
QY 434 ----- 433
Db 10710 EEAPAVEESPVEPPAGGVVWMPVSAKTSALDAQIGQLAAVAEDTVDPAVAARALV 10769
QY 434 ----- 433
Db 10770 DSRTAMEHRAVAGDSREALDALRMPEGLVRGTVDPGRAFVFPQGTQWAGMGABELL 10829
QY 434 ----- 433
Db 10830 DSSPEFAAAMAECEETALSPYVDWSLEAVVRQAPSAPTLRDVVVQVPTFAMVMSLAKVWQ 10889
QY 434 ----- 437
Db 10890 HHGITEAVIGHSGEIAAYVAGALTLDAAARVTLRSKSAIAHLAKGGMISLALSEE 10949
QY 438 ----- 437
Db 10950 ATRQRIENLHGLSIAAVNGPTATVVSQDPTQIQELAQCEADGIRARIIPVDYASHAHV 11009
QY 438 ----- 437
Db 11010 ETIENELADVLAGLSPTQVPPFFSTLEGTWITEPALDGGYVYRNLRHVRGAPAVETLA 11069
QY 438 ----- 437
Db 11070 TDEGTFHPIEVSHPVLTWLPDKVTGLATLRREDGQHRLTTSIAEAWANGLADWASL 11129
QY 438 ----- 441
Db 11130 LPATGALSPAVPLDPTYAFOHRSYMWISAPAGPGEAPHAHTASGREAVETGLAWPGAEDLD 11189
QY 442 ----- 444
Db 11190 EGRRSVLAWMVROAASVLRCDSPPEVPVDRPLREIGFDSLTVDFNRVNRNLTLGLQLP 11249
QY 445 ----- 451
Db 11250 PTVVFOHPTVALAERISDELAERNWAVAEPSDHEQAEKEKAAAPAGARSGADTGAGCM 11309
QY 452 ----- 451
Db 11310 FRALPQAVEDDRYGEFLDVLAEASAFRQFASPEACSERLDPVLLAGGTDRAEGRVVL 11369
QY 452 ----- 457
Db 11370 VGTGTAANGPHEFLRLSTSFQERDFLAVPLPGYGTGTGTALLPADLOTALDAQAR 11429
QY 458 ----- 457
Db 11430 AILRAAGDAPVLLGHSGGALLAHELAFRLERAHGAPPAGIVLVDPYPGHPQEPIEWSR 11489
QY 458 ----- 457

Db 11490 QLGEGLFAGEBPMDSARLLAMGRYARFLAGPRGRSSAPVLLVLRASEPLGDMQBERGDW 11549
QY 458 -----NSLW 461
Db 11550 RAHWDLPHVTADVDPGDHFTMRDHAPAVAEAVLSWLDATIEGEGAGKMTDRPLNVDSGLW 11609
QY 462 ----- 461
Db 11610 IRRFHPAPNSAVRLVCLPHAGGSASYFFRFSBELHPSVEALSVOYPCRQDRRAEPCLESV 11669
QY 462 -----WW 463
Db 11670 BELAEHVVAATEPFW 11684
RESULT 5
US-09-108-006C-1
; Sequence 1, Application US/09108006C
; Patent No. 6524613
; GENERAL INFORMATION:
; APPLICANT: Steer, Clifford J.
; Kren, Betsey T.
; Bandyopadhyay, Paramita
; Roy-Chowdhury, Jayanta
; TITLE OF INVENTION: Hepatocellular Chimeraplasty
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kimeragen, Inc.
; STREET: 300 Pheasant Run
; CITY: Newtown
; STATE: PA
; COUNTRY: USA
; ZIP: 18940
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/108,006C
; FILING DATE: 30-Jun-1992
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/054,288
; FILING DATE: 30-APR-1997
; APPLICATION NUMBER: 60/054,837
; FILING DATE: 03-AUG-1997
; APPLICATION NUMBER: 60/064,996
; FILING DATE: 10-NOV-1997
; APPLICATION NUMBER: 60/074,497
; FILING DATE: 12-FEB-1998
; APPLICATION NUMBER: PCT US 98/08834
; FILING DATE: 30-APR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Friebe, Thomas
; REGISTRATION NUMBER: 29258
; REFERENCE/DOCKET NUMBER: 7991-015-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-504-4444
; TELEFAX: 215-504-4545
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4563 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-108-006C-1

QY	Db	Best Local Similarity	4.7%; Pred. No. 0.0015;	Matches 213; Conservative 74; Mismatches 152; Indels 4091; Gaps 63;
12	QY	POMFVLISFLV- : : : : : : : : : : : : : : :	25	
6	Db	PALLALLAPALLLLLLAGARAEEMLENVSLVCPKDATRFXHLRKYTYNVEASSSGVP	65	
26	QY	-----SPINCKV- : : : : : : : : : : : : : : :	32	
66	Db	GTADRSATRINCKVEVPQLCSFILKTSQILKEVYGFNPEGKALLKTKNSEEPAAA	125	
33	QY	----- : : : : : : : : : : : : : : :	32	
126	Db	MSRYELKIAIPEGKQVFLYPEKDEPTYILNIRKGIISALLVPPETEAQVFLDVTVGN	185	
33	QY	----- : : : : : : : : : : : : : : :	32	
186	Db	CSTHFTVTRKGNVATEISTERDLGQCRPKPIRTGISPLALIKMTPLTLSSQSC	245	
33	QY	----- : : : : : : : : : : : : : : :	32	
246	Db	QYTLDAKRKHVAEAIKEQHLFLPPSYKNYGMVAQVTLKLEDTTPKINSRPPGEGTKK	305	
33	QY	----- : : : : : : : : : : : : : : :	32	
306	Db	MGLAFBSTKSTPPKQAEALVTQVELKKLITISEQNIQRANLKNLVTGLRGLSEAVTS	365	
33	QY	----- : : : : : : : : : : : : : : :	32	
366	Db	LLPOLIEVSPITLQALVOCQPOCSSTHILQWLKRVRHANPLLDVVTVLALIPESAQQ	425	
33	QY	----- : : : : : : : : : : : : : : :	32	
426	Db	LREIFNMARDQRSRATLYALSHAVNHYHKNPTGTQELLDIANYLMQIQDDCTGDEDT	485	
33	QY	----- : : : : : : : : : : : : : : :	32	
486	Db	YLILRVIGNMGTMEQLTPELKSSILKCVQSTKPSLMTQKAAIQALRMKBPDKQEVLL	545	
33	QY	----- : : : : : : : : : : : : : : :	32	
546	Db	QTFLLDASPDKRLAAYLMRSPSQADINKIVQLPWEQNEQVKNFVASHIANILNSEE	605	
33	QY	----- : : : : : : : : : : : : : : :	32	
606	Db	LDIQILKKLVKEVLESQPTVMDPRKFSRNQVLYKSVSIPSLDPASAKIEGNLIPDPNN	665	
33	QY	-----MSKALYNRPW- : : : : : : : : : : : : : : :	42	
666	Db	YLPKSSMLKTTLTAFGPASADLIEIGLKGKFEPTLEALFGKQGFPPDSVNKALY- -WV	722	
43	QY	-----RGLVLS- : : : : : : : : : : : : : : :	48	
723	Db	NGQVPDGVSKVLVDHFGYTKDKKHQDMVNGIMLSVSKLIDKLSKEVPEARAYLRILGE	782	
49	QY	KIGKYKLQQLK- : : : : : : : : : : : : : : :	65	
783	Db	ELGFASLHDQLGLKLLWGARTLQGIQPMIGEVIRKGSKNDFPLHYIPMENAFELPTGA	842	
66	QY	-----LETTISTTKYNS- : : : : : : : : : : : : : : :	77	
843	Db	GLQLOISSGVIAPGAKAGVKLEVANMQAELVAKPSVSVFVTNMGIIIPDFARSGVQMN	902	
78	QY	-----KOPVKNL- : : : : : : : : : : : : : : :	89	
903	Db	TNPFHESGLEAHVALKPKGLKPIIPSPKRPVKLLSGGNTLHLVSTTKTEVIPPLIENROS	962	
90	QY	-----PP- : : : : : : : : : : : : : : :	98	
963	Db	WSVCKQVFFGLNYCTSGAYSNASSTDSASYPLGDTFLLELRPTGEIEIQYSVSATYEL	1022	
99	QY	----- : : : : : : : : : : : : : : :	98	

```
QY 216 ----- 215
Db 2163 AKINFNEKLSQLOQYMIQFDYIKDSYDLHDLKIAIANIIDEIIEKLSDEHYHVRNL 2222
QY 216 ----- 215
Db 2223 VKTIHDLHLFIENIDFNKSGSSTASIQNVDTKYQIRIQIQEQKLOQKRIHQIHDIOHLA 2282
QY 216 ----- QRNCTHSFYLVNAMSRL ----- FRVPKYING----- 242
Db 2283 GKLOQHEADVRVLLDQLTGISFERINDVLEHVHVFVNILIGDFEVAEKINAPRAKVH 2342
QY 243 ----- TKLQNTMKLKR--KOAPVKQOFEK----- 265
Db 2343 ELIEREYVDOQIQVLMDKVELAHQYKELTIQKLSNVLQOVKIKDYFEKLVGFIIDAVK 2402
QY 266 ----- 265
Db 2403 KLNELSKTFIEDNVNKFLOMLIKKLSFDYHQFVDETNDKIREVTQRLNGEIQALELPQK 2462
QY 266 ----- KAK----- 268
Db 2463 AEALKFLBETKATVAVYLESQDTKITLIINWLQEALSASLAHMKAKERETLEDRDR 2522
QY 269 ----- KTGSTTPYFS-YTTSAALNVTTNV-TYSITTAARRVSTSTI 308
Db 2523 MYQMDIQBELQRYLSLVGVYSTLVYISDWMTLAAKNLTDFAEQYISIQDWAKRMKALV- 2581
QY 309 AYRPDSSF--KXSIWATO----- 324
Db 2582 -----EQGFTVPEIKITLGTMPAPEVSLQALQKATFQTPDIVPLTDLRIPSQVINFKDLK 2637
QY 325 -----LRDL- 328
Db 2638 NIKIPRSTPEITLNTHIPSTIDFVEMKVKIIRTIDQMLNSLQHPVPIYLRDLK 2697
QY 329 ----- 328
Db 2698 VEDIPLARITLPDFRLPEIAPEFIPTLNLNDQVDPDLHIPFOLPHISHTIEVTFGK 2757
QY 329 ----- 328
Db 2758 LYSILKIQSPLFLTDANADIGNTTSANEBAGIAASITAKGESKLEVLNDFDQANAQSNP 2817
QY 329 ----- 328
Db 2818 KINPLAKESVKPSSKYLARTEHGSEMLFFPGNAIEGKSNTVASLHTEKNTLELSNGVIVKI 2877
QY 329 -----ATWVYTT----- 335
Db 2878 NNQLTDSNTKYPHKLNI PKLPSSQADLRNEIKTLKAGHIAWTSSGKSGMKWACPRFS 2937
QY 336 -----LRYRON----- 341
Db 2938 DEGTESQISFTTIEGELTSGLSNKINSKHLRVNQNLVYESGSLNFSKLEIQSDVSQHV 2997
QY 342 -----PF-CEPSRNR----- 350
Db 2998 GHSVLTKAGMALFGEKGAEFTGRHDAHLNGKVI GTLKNLSLFFSAQPFETASTNNEGNLK 3057
QY 351 -----TAVSEFMKN-----THVLI 364
Db 3058 VRPPLALTGKIDFLNNVALFLSPSAQAASQVQSAREFNQYKYNQNFSGAGNENIMEARVGI 3117
QY 365 RNET-----PYTIYGT----- 375
Db 3118 NGEANLDFLNIPLTIPEMLPYTIITTPPLKDFSLWEKTLKEFLTKTTKQSFDSVKAQY 3177
QY 376 -----LDMSLYNET----- 386
Db 3178 KQKXHRHSITNPLAVLCBFISQISKDFDRHFEKQNRNALDFVTKSYNETKIPDKYKAEK 3237
```

```
QY 387 ----- 386
Db 3238 SHDELPRTFQIPGYTVPVVNVVEVSPFTIEMSAFGYVFPKAVSMPFSILGSDVRVPSYTL 3297
QY 387 ----- 386
Db 3298 ILPSLEPLVHLVPRNUKLSLPDPKELCTISHIFIPAMGNITYDFSPKSSVITLNTNAELF 3357
QY 387 -----MFVE---NKTAS-- 395
Db 3358 NOSDIVAHLSSSSSVIDALQYKLEGTTRLTRKRGKGLATALSLSNKFVEGSHNSTVSLT 3417
QY 396 -----DSNKTTP-SPSMGFQTF----- 413
Db 3418 TKRMEVSVATTTKAQPIILRMNFQBELNGNTKSKPTVSSSMFEKYDFNSMLYSTAKGAV 3477
QY 414 ----- 413
Db 3478 DHKLSLESLSYFSIESSTKGDVKSLSREYSGTTIASEANTYLSKSTRSSVKLOQTSK 3537
QY 414 ID----- 415
Db 3538 IDDNWLEVKENFAGEATLQRIYSLWEHSTKNHLQLEGLFPTNGEHTSKATLELSPWOMS 3597
QY 416 ----- 415
Db 3598 ALVQVHASQPSSSHFDPPDIGQEVNANTKNQIRKWEVRIHSGSFQSQVELSNDQEKA 3657
QY 416 -----PLWDL----- 421
Db 3658 HLDIAGSLEGLHURLFNKIILPVYDKSLWDFLKDVTTSIGRRQHLRVSTAFVYTKPNGY 3717
QY 422 ----- 421
Db 3718 SFSIPVKVLADKFIIPGLKLNLSVLMPTFHVPTDLQVPCKLDFREIYIYKKLRTS 3777
QY 422 -----DSLLFLDEI----- 430
Db 3778 SPALNPLTLPEVKFPEDVLTQYQPEDSLIPFEITVPESQTVSQFTLPKSVSDGIAA 3837
QY 431 ----- 430
Db 3838 LQNAVNAKIADPELPTIIVPEQTIEIPSIKFSVPAGIAIPSFQALTARFVDSVPYNAT 3897
QY 431 -----RNFS-- 434
Db 3898 WSASLKNKADYVETVLDSCTSVQFLEYELNVGLTHKIEDGTFLASKTKGTFAHRDFAE 3957
QY 435 -----LRSPT----- 439
Db 3958 YEEDGKYEGLOEWEGKAHLNISKAPFTDLHLRYQDKKGIKSTSAASPAVGTVMQMDDED 4017
QY 440 ----- 439
Db 4018 DFSKNFYFSSPOSSPKKLTIFKTELVRRESDEETOIKNVWEEAASGLLTSLKONVPA 4077
QY 440 ---YVN----- 442
Db 4078 TGVLYDYVKNKYHEHTGLTLREVSSKLRNLODHAEVWYQGAIREIDIDERFQKGASGT 4137
QY 443 ----- 442
Db 4138 TGTQEWKDKAQNLYQELLITQEQASFQGLKDNVDFGLVRVTQEFHMKVKHLIDSIDFL 4197
QY 443 ----- 442
Db 4198 NPPRFQPGKPGIYTREELCTMFIREVGTLSQVSVKNGSEILPSYFODLVITLPFEL 4257
QY 443 ----- 442
Db 4258 RKHKLIDVISMYRELLKDLKSAQEVFKAIQSLKTTTEVLRLNQLDLQFIFOLLIEDNIKQL 4317
QY 443 ----- 442
```

Db 4318 KEMFTYLYNQDEINTFNDYIPYVFKLKNCLNHLKFNFIQNELQASQEIQQI 4377
 QY 443 -----LTPP----- 442
 Db 4378 HOYIMALREBYDPSIVGWTVKYVELEBKIVSLIKNLLVALKDPHSEYIIVSASNFTSOLS 4437
 QY 443 -----LTPP----- 446
 Db 4438 SQVEQFLHRNIQVLSILTPDGGKGKIEAELSATAQEIISQAIATKLIISDYHQQFRI 4497
 QY 447 -----EHRVAVNLSTSN 458
 Db 4498 KLQDFSDQLSDYVEKFAESKRLIDLSIQN 4527

RESULT 6
 US-09-538-092-842
 ; Sequence 842, Application US/09538092
 ; Patent No. 6753314
 ; GENERAL INFORMATION:
 ; APPLICANT: Giot, Loic
 ; APPLICANT: Mansfield, Traci A.
 ; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
 ; FILE REFERENCE: 15966-542
 ; CURRENT APPLICATION NUMBER: US/09/538,092
 ; CURRENT FILING DATE: 2000-03-29
 ; PRIOR FILING DATE: 1999-04-01
 ; PRIOR APPLICATION NUMBER: 60/127,352
 ; PRIOR FILING DATE: 1999-04-01
 ; PRIOR APPLICATION NUMBER: 60/178,965
 ; NUMBER OF SEQ ID NOS: 1387
 ; SOFTWARE: CuratPatSeqFormer Version 0.9
 ; SEQ ID NO 842
 ; LENGTH: 4563
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (0)...(0)
 ; OTHER INFORMATION: Polypeptide Accession Number P04114
 US-09-538-092-842

Query Match 18.3%; Score 449; DB 4; Length 4563;
 Best Local Similarity 4.6%; Pred. No. 0.0015;
 Matches 209; Conservative 74; Mismatches 151; Indels 4101; Gaps 61;
 QY 12 PKMFVLISFLAV----- 25
 Db 6 PALALALFALLLLLAGARAEEMLENVSLVCPKDATRFKHLRYTYNREABSSGVP 65
 QY 26 -----SFINCKV----- 32
 Db 66 GTADRSATRNCKVEVPQLCSFILKTSQCTLKEVYGFNPEGKALLKTKNSBEFAAA 125
 QY 33 ----- 32
 Db 126 MSRYELKLAPEGKQVFLYPEKDEPTYILNIKRGIIISALLVPPETBEAKQVFLDTVYGN 185
 QY 33 ----- 32
 Db 186 CSTHTVTKRGNVATEISTERDLGQDRFKPIRTGISPLALIKGMRPLSTLISSQSC 245
 QY 33 ----- 32
 Db 246 QVTLDAKRKHVAEAIKCEQHLFLPFSYNNKYGVAQVOTLKLDDTPKINSRFFGEGTKK 305
 QY 33 ----- 32
 Db 306 MGLAFESTKTSPPKQAEAVLKTQELKKLTISEQNIORANLFNKLVTELRLGLSDEAVTS 365
 QY 33 ----- 32

Db 366 LLPQLIEVSSPITLQALVOCGQPCSTHILQWLKRVPANPLLDVVYLVVALIPEPSAQO 425
 QY 33 ----- 32
 Db 426 LREIFNMARDORSRATLYALSHAVNNYHKTNPGTQELLDIANYLMEIQIDDCDGEDYT 485
 QY 33 ----- 32
 Db 486 YLILRVIGNMGQTMEOITPELAKSSILKCVQSTKPSLMIKAAIQALRKMEPKDKQDEVLL 545
 QY 33 ----- 32
 Db 546 QTFLDASPCDKRLAAYLMLRSPSQADINKIVILPWEQNEQVKNFVASHIANILNSE 605
 QY 33 ----- 32
 Db 606 LDIQDLKLIVKEALKESQLPTVMDPRKFRNRYLYKVSLSPLSPASAKIEGNLIFDPNN 665
 QY 33 -----MSKALYNRPW- 42
 Db 666 YLPKESMLKTTLTAFGPASADLIEIGLECKGPEPTLEALFGKQGFPPDSVNKALY--WV 722
 QY 43 -----RGLVLS----- 48
 Db 723 NGQVPDGVSKVLVDHFGYTKDDKHEQDMVNGIMLSVEKLIKDLKSKVEPEARAYLRLGE 782
 QY 49 ----- 48
 Db 783 ELGFASLHDLQLLGLKLLMGARTLOQIPQIMGIVIRKSGKNDFFLHYIPMENAFELPTGA 842
 QY 49 ----- 48
 Db 843 GLQLQISSSGVIAPGAKAGVKLEVANMQAELVAKPSVSVFVTNMGIIPDFARSGVQNN 902
 QY 49 -----KIGKYK----- 54
 Db 903 TNFFHESGLEAHVALKAGKLFIIIPSPKRPVKLLSGGNTLHLVSTTKTEVIPPLIENROS 962
 QY 55 ----- 54
 Db 963 WSVCKQVFPGLNYCTSGAYSNASTSDSASYPLTGDTRELELRPTGEIEQVSVSATVEL 1022
 QY 55 -----LDQLKL-----BILRQLETTISTKYNVSKQPVKNLTMTTE--FPQY- 93
 Db 1023 QREDRALVTTLKFVTOAGAKQTEATWTFKN-----RQSMTLSEVOIPDPDVLGTIL 1077
 QY 94 -----YILAGPIQNYSIYTL----- 108
 Db 1078 RVNDESTEGKTSYRLTLDIONKKITEVALMGHLSCDTKEERKIKGVISIPRLQABARSEI 1137
 QY 109 ----- 108
 Db 1138 LAHWSPAKLLQMDSSATAYGSTVSKRVAMHYDEKIEFENWTGNVDTKMTSNFPVDL 1197
 QY 109 -----WF----- 110
 Db 1198 SDYPKSLHWANRLIDHRVPETDMTFRHVSGKLIAMSSWLQKASGSLPYTQTLQDHLNS 1257
 QY 111 -----DFVSTQ----- 118
 Db 1258 LKEFNLQNMGLPDPHIPENLFLKSDGRVKYTLNKNLSKIEIPLPPGGKSRDLKMLETVR 1317
 QY 119 KPAKY-----VYSQ-YNHNTA----- 132
 Db 1318 TPALHFKSVGFHLPSREPVFTTIPKLYQLQVPLGLVLDLSTNVYSLNKNWSYSGGN 1377
 QY 133 -----KTITP----- 137
 Db 1378 TSDHFSLRARYHMKADSVDDLSTNVQSGGETTYDHKNTFTLSCDGLSRHKFLDSNIFK 1437
 QY 138 ----- 137
 Db 1438 SHVEKLGNNPVSKGLLIFDASSWGPQMSASVHLDSKKKQHLFVKEVKIDQPRVSSPYA 1497

QY 138 -----RPPCGRVP-----SNTCLSEM-----154
DB 1498 KGTGGLSCORDPNTGRNGESNLFNSYLGQNTQITGRYEDGTLSTSTSDLSQGIKN 1557
QY 155 -----LNVSERN-----161
DB 1558 TASLKYENVELTKSDTNGKYKNFATSNKMDMTFSKONALLRSEYQADYESLRFFSLSG 1617
QY 162 -----DTGB 165
DB 1618 SLNSHGLELNADILGTDKINSAGAKATLRIGQDGISTATSATTNKLKCSLLVLENELNAELGL 1677
QY 166 QCGG-----NF-----171
DB 1678 SCASMKLTNGRPREHNAKFLDGLKAALTELSGAYQAMILGVDGSKNIFNFKVSOEGLK 1737
QY 172 -----TTF 174
DB 1738 LSNDDMGSAEMKFDHTNSLNIAGLSLDFSSKLDNIYSSDKFYKQTNVLQLPYSLVTL 1797
QY 175 N-----PM-----177
DB 1798 NSDLKYNALDLTNGKLRLEPLKLVAGNLKGAYQNEIKHIYAISSAALSASYKADTVA 1857
QY 178 -----FNVPR-----183
DB 1858 KVGQVFSHRLNTDIAGLASAIDMTNYSNLSLHFSNVFRSVMAPTMTIDAHTNGNGL 1917
QY 184 -----183
DB 1918 ALWGEHTGOLYSKFLKABPLAFTFHDYKGSTSHHLVSRKSISSAALHKVSALLTPAQ 1977
QY 184 -----WTKLVG-----PTKV-----195
DB 1978 TGTWKLKTOFNNEYSQDLDAYNTKKIGIVELTGRTLADLTLLDSPIKVPLLSEPINII 2037
QY 196 -----NVDSTI-----YF-----204
DB 2038 DALEMRDAVEKPOEFTIVAPVKYDKQVHSINLPFFTELQVEYFERNROTIIVVENVQR 2097
QY 205 -----LG-----LTALLRYA-----215
DB 2098 NLKHINIDQVRKYRAALAGKLPOANDYLSNFNWERQVSHAKEKLTALTCKYRITENDIQ 2157
QY 216 -----215
DB 2158 IALDDAKINFNEKLSOLOQTMIOFDQYIKDSYDLHLDKIAIAIIDEIIEKLKSLDEHYH 2217
QY 216 -----215
DB 2218 IRVNLVKTIDHLFTENIDFNKSGSSTASWIONVDTKYQIRIQIQEKLQQLKRHIQND 2277
QY 216 -----QRNCTHSPYLVNMSRNL-----FRVPKYING- 242
DB 2278 IQHLAGLKKQHIBAUDRVLLDQGTGISFRINDVLEHVKHFVINLIGDVEFAEKINAF 2337
QY 243 -----TKLNTMRKLKR--KOAPVKEOPEK-----265
DB 2338 RAKVHELIEREYVDQIQVLMKDLVELTHQYKLUKTIQKLSNVLOQVKIKDYFEKLVGFI 2397
QY 266 -----265
DB 2398 DDAVKKLNLSFKTFIEDYNKPLDMLIKKLKSFYDHFVDETNDKIREVTQRLNGEIQAL 2457
QY 266 -----KAK-----268
DB 2458 ELPQKAEALKFLEETKATVAVYLESIQDTKITLIINWLQEALESASLAHMAKAFRETLE 2517
QY 269 -----KTOSTTPPYES-YTTSALNVTTNV-TYSITTTAARV 303
DB 2518 DTRDRMYQMDIQELOYRLSLVGQVYSTLVTYISDWTIAARNLDTFASQYSIQDWAKRM 2577

QY 304 STSTIAYRPPDSF-----MKSIMATQ-----324
DB 2578 KALV-----EQGFTVPEIKTILGTWPAFEVSLQALOKATQTFDFIVPLTDLRIPSQIN 2632
QY 325 -----324
DB 2633 FKDLKNKIPSRFSPFTILNTFHIPSTIDFVEMKVKLIIRTIQONQSELOWPVPDIY 2692
QY 325 LRDL-----328
DB 2693 LRDLKVEDIPLARITLDPDRLEPAIPEFIPTLNLDFQVPLDHIPEFQPHISHTEIV 2752
QY 329 -----328
DB 2753 PTFGKLYSILKIQSPLFTLDANADINGTTSANEAGIAASITAKGBSKLEVLNDFPOANA 2812
QY 329 -----328
DB 2813 QLSNPKINPLALKESVKFSSKYLRTHEGSEMLFFGNAIEGKSNVTASLHTEKNTLELSNG 2872
QY 329 -----ATWVYT 334
DB 2873 VIVKINNQLTLDSTNKYPHKLNIPLKDFSSQADLRNEIKTLKAGHIAWTSSGKGSWKA 2932
QY 335 T-----LRYRON-----341
DB 2933 CPRFSDGTHESQISFTIEGPLTSFELSNNKINSKHLRVNQNLVYESGSLNFSKLEIQSQV 2992
QY 342 -----PF-CBPSRNR 350
DB 2993 DSQHVHVSULTAKGMALFQEGKAEFTGRHDAHLNGKIVIGTLKNSLFFSQPFEITASTNN 3052
QY 351 -----TAVSEFMKN-----359
DB 3053 EGNLKVREPLRLTGKIDFLNNYALFLPSAQASQVSAFENQYKYNQNSAGNNENIME 3112
QY 360 THVLIRNET-----PYTIYGT-----375
DB 3113 AHVINGEANLDFNLPIPLTIPEMRLPYTIITPPKDFSLWEKTLGKBEFLKTTKQSFDSL 3172
QY 376 -----LDMSSLYNET-----386
DB 3173 VKAQYKKNKGRHSITNPLAVLCEFSQISKSPDRHPEKRNALDPVTKSYNETKIKPKDK 3232
QY 387 -----386
DB 3233 YKAEKSHDELPTFQIPGYTVPVNVVESPFTIEMSAFGYVFPKAVSMPSPSLGSDVRV 3292
QY 387 -----386
DB 3293 PSYTLILPSLELPVLHVPRNLKSLPHPKELCTISHIPAMGNITYDFSKSSVITLNT 3352
QY 387 -----MPVB---NK 392
DB 3353 NAEFLNQSDIVAHLLSSSSSVIDALQYKLEGTTLTRKGLKLATALSLSNKFVEGSHNS 3412
QY 393 TAS-----DSNKTPT-SPSMGFORTF-----413
DB 3413 TVSLTTKNMEVSAKTKABEIPILRMNFKQELNGNTKSKPTVSSSMEPKYDFDNSSMLYST 3472
QY 414 -----413
DB 3473 AKGAVDHKLSELSLTSYFSIESSTKGDVKGSVLSREYSGTIASEANTYLSKSTRSSVKL 3532
QY 414 -----ID-----415
DB 3533 QGTSKIDDIWNLEVKENFAGRATLORIYSLWEHSTKNHLOLEGLPTNGEHTSKATLELS 3592
QY 416 -----415
DB 3593 PHQMSALVOVHASQPSSEPHDFDLGQEVNANTKNQIRWKNVEPRIHSGSPSQSVELSN 3652
QY 416 -----PLWDYL-----421

Db 3653 DQKALHDIAGSLGHLRFLKNIILPVYDKSLMDFLKLDTVTSIGRRQHLRVSTAFVYTK 3712
QY 422 -----
Db 3713 NPNQYSFIPVKVLADKFIITPGLKLDNLNSVLVMPFTHVFTDLQVPCKLDFREIQIYK 3772
QY 422 -----
Db 3773 KLRTSSPALNLPLPEVKPFPEVDVLTYSQPEDSLIPFFFIITVPESQLTVSQTLPSKVS 3832
QY 431 -----
Db 3833 DGIAALDLNAVANKIADFELPTIIVPEQTIEIPSIFSPAGIVIPSFQALTARFEVDSP 3892
QY 431 -----
Db 3893 VYNATWSASUKKADYVETVLDSTCSVQFLEYELNVGLTHKIEDGTLASKTKGTLAHR 3952
QY 432 NFS-----
Db 3953 DFLSAEVEDCKFGLQEWGKAHLNLIKSPAFDHLRYQDKKIGISTASAAPAVGTVMGD 4012
QY 440 -----
Db 4013 MDEDDDFSKWPFYVSPQSPDKLTIFKTELVRSEDEETQIKVNWEEAASGLTSLKD 4072
QY 440 -----
Db 4073 NVPKATGVLYDYNNKYHWEHTGLTREVSSKLRLNQNNAEWYOGAIRQIDIDVRFOK 4132
QY 443 -----
Db 4133 AASGTTGTQEWKDKAQLYQELLTQEQASFOGLKDNVDFGLRVTKQHKMKVHLIDS 4192
QY 443 -----
Db 4193 LIDFLNPPRPFPKPGIYTBELCTMFIREVGTLSQVYKVHNGSEILFSYFDLVIT 4252
QY 443 -----
Db 4253 LPPELRKHLIDVISMVRELLKDSKEAOEVFKAIOSLKTTEVLRLNQLLOFIFOLIED 4312
QY 443 -----
Db 4313 NIKQLKEMKFTYLVINQIDEINTIPNDYIPYVFKLKENLCLNLHKFNEFIQNELQESQ 4372
QY 443 -----
Db 4373 ELQIHHQYIMALREBYFDPISVGTWKYVELEBKIVLKNLILVALKDFHSEYIVSASNF 4432
QY 443 -----
Db 4433 TSOLSSQVQFLHRNTQEVLSILTDGKGKEKIAELSATAQBIKSOAIATKKIISDYH 4492
QY 447 -----
Db 4493 QOPRYKLQDPSDQSDYVEKFAESKRLIDLISQN 4527

RESULT 7

US-09-180-422B-27
; Sequence 27, Application US/09180422B
; Patent No. 6444644
; GENERAL INFORMATION:
; APPLICANT: BRUCKDOERFER, KARL R
; ETTLEAIE, CAMILLE
; TITLE OF INVENTION: ANTICOAGULANT PEPTIDE FRAGMENTS DERIVED
; FROM APOLIPROTEIN B-100
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE, P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR

CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/180,422B
FILING DATE: 07-Dec-1998
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: SADOFF, B.J.
REGISTRATION NUMBER: 36663
REFERENCE/DOCKET NUMBER: 117-268
TELECOMMUNICATION INFORMATION:
TELEPHONE: 7038164000
TELEFAX: 7038164100
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 4536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-180-422B-27
Query Match 18.0%; Score 440; DB 4; Length 4536;
Best Local Similarity 4.8%; Pred. No. 0.003;
Matches 213; Conservative 76; Mismatches 156; Indels 3970; Gaps 64;
QY 1 MGKKE-----5
Db 99 MSRYELKLAIPGKQVFLYPEKDEPTYILNKRGIISALLVPPEEAKQVLFDTVYGN 158
QY 6 -----5
Db 159 CSTHTVTKRKNVATEISTERDLGQCDRFKPIRTGISPLALIKGWRPLSTLISSQSC 218
QY 6 -----MMVRDVPK-----13
Db 219 QYTLDAKRHVAEAIKESQHLFLPFSYKNKYGMVAQVQTQTKLEDTPKINSRPFGEGTK 278
QY 14 -----MP-----15
Db 279 MGLAFESTKSTSPPKQAEAVLTKLQELKLTISEQNIQRANLFNKLVTRELGLSDEAVTS 338
QY 16 -----VLISIFLIVSEI-----28
Db 339 LLPQLIEVSSPITLQALVQCGPQCSSTHLOWLKRVRHANPLLDIVTVYLVALLIPEPSAQ 398
QY 29 -----28
Db 399 LREIFNWARDQSRATLYALSHAVNNYHKTNPGTQELLDIANYLMEQIQDDCTGDEYDT 458
QY 29 -----28
Db 459 YLILRVIGNMGQTMELTPELKSSILKCVQSTKPSLMIQAAIQALRKMPEKDKQEVLL 518
QY 29 -----28
Db 519 QTFLLDDSPGDKRLAAYLMLMRSPSQADINKIVQLPWEQNEQVKNFVASHIANILNSEE 578
QY 29 -----NCKV-----32
Db 579 LDIQDLKKLVKEVKESQLPTVMDPKFSRVQYKSVSLPSLDPSAKIEGNLIFDPNN 638
QY 33 -----MSKALYNRPW-42
Db 639 YLPKESMLKTTLTAFGFASADLIEIGLEGKGFEPTEALFGKQGFDPDSVKNALY---WV 695

```
QY 43 -----RGLVLS----- 48
Db 696 NGQVPDGVSKVLVDHFGYTKDDRHEQDMVNGIMLSVEKLIKDLKSVEPEARAYLRILGE 755
QY 49 ----- 48
Db 756 ELGFASLHDLQLLKGKLLMGARTLOGIPOMIGEVIRKSGKNDPFLHYIPMENAFELPTGA 815
QY 49 ----- 48
Db 816 GLQLOISSSGVIAPGAKAGVKLEVANMQAELVAKPSVSEFVTNMGIILIPDPARSQVQWN 875
QY 49 -----KIGKYK----- 54
Db 876 TNFFHESGLEAHVALKAGKLFPIIPSPKRPVKLLSGNTLHLVSTTKTEVIPPLIENRQS 935
QY 55 ----- 54
Db 936 WSVCKQVFPGLNYCTSGAYSNASSTDSASYPTGTTRLELRPTGEIEQVSVSATYEL 995
QY 55 -----LDQLKL-----EILRQLETTISTKYNVSKQPVKNLTWNTB--FPQY----- 93
Db 996 QREDRALVTLKFVTOAGKAKQTEATWTFKN-----RQSMILSSEVOIPDFDVLGTIL 1050
QY 94 -----YILAGPIQNSITYL----- 108
Db 1051 RVNDESTEGKTSYRLTLDIQNKKITEVALMGHLSGCDTKEERKIKGVISIPRLQAEARSEI 1110
QY 109 ----- 108
Db 1111 LAHWSPAKLLQMDSSATAGYSTVSKRVAMHYDEEKIBFEWNTGTNVDTKGMTSNPPVDL 1170
QY 109 -----WF----- 110
Db 1171 SDYPKSLHMYANLLDHRVPQDMDTFRHVGSKLIVAMSSWLQKASGSLPYTQTLQHLNS 1230
QY 111 -----DFYSTQ-----LR 118
Db 1231 LKEFNLQNMGLPDFHIPENFLKSDGRVKYTLNKSLEKTEIPLPFGGKSRDLKMLETVR 1290
QY 119 KPACY-----VYSQ--YNHTA----- 132
Db 1291 TPALHFKSVGFHLPSREFQVPTFTIPLKYQLQVPLLVGLDLSNTVYSLNLYNSASYSGGN 1350
QY 133 -----KITP----- 137
Db 1351 TSTDHFLSLRHYMKADSVVDLLSYNVQSGSETTYDHKFTLSCDGLSRHKFELDSNIKF 1410
QY 138 ----- 137
Db 1411 SHVEKLGNNPVSKGLLIPDASSSWGPMQMSVHLDSSKKQHLFPVKEVKIDGQFRVSSFYA 1470
QY 138 -----RPPPCGRVP-----SMTCLSEM----- 154
Db 1471 KGTYGLSCORDPNTGRNGESNLRFNSSYLQGTNQITGRYEDGTLSTSTSLQSGIKN 1530
QY 155 -----LNVSKRN----- 161
Db 1531 TASLKYENYELTUKSDTNKYNKPNFATSNKMDTFFSKQALLRSEYQADYESLRPFFSLSG 1590
QY 162 -----DTGE 165
Db 1591 SLNSHGLELNADILGTDKINGGAHKATLRIGQDGI STSATTNLKSLVLLENELNAELGL 1650
QY 166 QCGG-----NF----- 171
Db 1651 SGASMKLTJNGRPREHNAKFLSDGKAALTELSIGSAYQAMILGVDSKNIFNFKVSOEGLK 1710
QY 172 -----TTF 174
Db 1711 LSNMGMGSAEMKFDHTNSLNIAGLSLDFSSKLDNIYSSDKFYKQTVNIQLQPSLVLTTL 1770
```

```
QY 175 N-----PM----- 177
Db 1771 NSDLKYNALDLTNNGKRLRLEPLKHLVAGNLKAGAYQNEIKHIYAISSAALSASYKADTV 1830
QY 178 -----FENVPR----- 183
Db 1831 KVQGVFESHRLNTDITAGLASAIDMTSTNYNSDSLHSFNVSFVMAFPFTMTIDHTNGNGL 1890
QY 184 ----- 183
Db 1891 ALWGEHTGQYSKFLKAEPLAFTSHLDYKSGTSHHLVSRKSISSAALEHKVSALLTPAEQ 1950
QY 184 -----WNTKLYVG-----PTKV----- 195
Db 1951 TGTWKLKTQFNNEYSQDLDAYNTKDKIGBELTGRTLADLTLLDSPIKVPLLSEPINII 2010
QY 196 -----NVDSTI-----YF----- 204
Db 2011 DALEMRDAVEKPOEFTIVAFVKYDKNQDVHSINLPFFETLOEYFERNQRTIIVLENVQR 2070
QY 205 -----LG-----LTALLRYA----- 215
Db 2071 NLKHINIDQFVKYRAALGKLQQANDYLNSFNWERQVSHAKEKLTALTUKYRITENDIQ 2130
QY 216 ----- 215
Db 2131 IALDDAKINFNEKLSOLOTYMIQDQYIKDSYDLHDLKIAIANIIDEIIEKLSLDEHYH 2190
QY 216 ----- 215
Db 2191 IRVNLVKTIDHLHFIENIDFNKSSGSTASWTONVTKYQIRIQIOEKLQOLKRHTQND 2250
QY 216 -----QRNCTSHFYLVNAMSRL-----FRPKYING----- 242
Db 2251 IQHLAKLQHIIEADVRVLLDQGTTSIFERINDVLEHVHGFVNILGDFEVAEKINAF 2310
QY 243 -----TKLKNTRMKLKR--KOAPVKEQPEK----- 265
Db 2311 RAKVHELIEREVDQIQIQLMDKCLVELAHQYKLETKIQKLSNVLOQVKIKDYPEKLVGFI 2370
QY 266 ----- 265
Db 2371 DDAVKKLNELSPKTFIEDVNKFLDMLIKKLKSDYHQFVDETNDKIREVTQRLNGEIQAL 2430
QY 266 -----KAK----- 268
Db 2431 ELPOKAEALKFLEETKATVAVVLESLODKITLIINWLQEALSSASLAHMAKAFRETLE 2490
QY 269 -----KTQSTTPPYFS-YTTSAAALNVTTNV-TYSITTAARRV 303
Db 2491 DTRDRMYQMDIQOELQRYLSLVGVYVTLVTYISDMWTLAAKNLTDPAEQYSIQDWAKRM 2550
QY 304 STSTIAYRPDSSF-----MKSINATQ----- 324
Db 2551 KALV-----EQGTVPETKITLGTWPAFVLSQALQKATQTPDFIVPLTDLRIPVQIN 2605
QY 325 ----- 324
Db 2606 FKDLKNIKIPRSPTPEFTILNTFHPSTIDFVEMKVKIIRTIDQMLNSELQWPVDDIY 2665
QY 325 LRDL----- 328
Db 2666 LRDLKVEDIPLARITLPDFRPLPEIAIPEIPIITLNLNDFQVPLHPIBFLQPHISHTIEV 2725
QY 329 ----- 328
Db 2726 PTFGKLYSLTKIOSPLFTLDANADICNGTTSANEAGIAASITAKGESKLEVLNDFQANA 2785
QY 329 ----- 328
Db 2786 QLSNPKINPLALKESVFKSSKYLRTHEGSEMFLFPFGNAIEGKSNVTASLHTEKNTLELSNG 2845
QY 329 -----ATWVYT 334
```


; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4536 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-09-079-030-1

Query Match

Best Local Similarity 4.8%; Pred. No. 0.003;

Matches 213; Conservative 76; Mismatches 156; Indels 3970; Gaps 64;

18.0%; Score 440; DB 4; Length 4536;

```
QY 1 MGRKE----- 5
DB 99 MSRYELKLAPEGKQVFLYPEKDEPTYILNKRGIISALLVPPETBEAKQVFLDTPYGN 158
QY 6 ----- 5
DB 159 CSHPTVTRKGNVATEISTERDLGQDRPKPIRTGISPLALIKGMRPLSTLISSQSC 218
QY 6 -----MMVRDPK----- 13
DB 219 QYTLDAKRKHVAEAIKCEQHLFLPFSYNNKYGVAQVOTLKLDTPKINSRFFGEGTKK 278
QY 14 -----MF----- 15
DB 279 MGLAFESTKTSPPKQAEAVLKTBLKLTISEQNIQRANLFNKLVTBLRGLSDEAVTS 338
QY 16 -----VLISIFLLVSP----- 28
DB 339 LLPQLIEVSPITLQALVQCGQPCSTHILQWLKRVHNPFLLDVTVYLVALIPEPSAQO 398
QY 29 ----- 28
DB 399 LREIFNWARDQSRATLYALSHAVNVYKTNPTGTQELLDIANYLMEIQDCTGDEDYT 458
QY 29 ----- 28
DB 459 YLILRVIGNMQTMEQLTPELKSSILKCVQSTKPSLMIKAAIQAALRKMPEKDKQDEVLL 518
QY 29 ----- 28
DB 519 QTFLDASPGDKRLAAYLMLMRSPSQADINKIVQILPWEQNEQKRVFVASHIANILNSEE 578
QY 29 -----NCKV----- 32
DB 579 LDIQDLKLVKEALKESQLPTVMDFRKSRNYQLYKSVSLPSLDPASAKIEGNLIFDPNN 638
QY 33 -----MSKALYNRPW- 42
DB 639 YLPKESMLKTLTAFGFASADLIEIGLEGKGFPTLEALFGKQFPFDSVNKALY---WV 695
QY 43 -----RGLVLS----- 48
DB 696 NGQVPDGVSKVLVDHFGYTKDDKHEQDMVNGIMLSVEKLIKDLKSKEVPEARAYLILGE 755
QY 49 ----- 48
DB 756 ELGFASLHDLQLLGLLLMGARTLQGI PQMIGEVIRKGSKNDFFLHYIFWENAFELPTGA 815
QY 49 ----- 48
DB 816 GLQIQISSGVIAPGAKAGVKLEAVANQAEVLVAKPSVSEVFTNMGIIPDFARSQVMN 875
QY 49 -----KIGVK----- 54
DB 876 TNFFHESGLEAHVALKAGLKFIPSPKRPVKLLSGNTHLVSTTKTEVIPPILIENROS 935
QY 55 ----- 54
DB 936 WSVCKQVPGFLNYCTSGAVSNASSTDSASYPTGTDRLELRPTGEIEQVSVSATYEL 995
QY 55 -----LDQLKL-----EILRQLETTISTKYNVSKQFVKNLWNTE--PFOY----- 93
```

```
DB 996 QREDRALVDTLAKFVTOAEKAKQTEATMTFKYN-----ROSMTLSSSEVQIPDFDVLDTGL 1050
QY 94 -----YILAGPIQNTSYITV----- 108
DB 1051 RVNDESTEGKTSYRLTLDIQNKKIIEVALMGHLSCDTKERKIKGVISIPRLQAEARSEI 1110
QY 109 ----- 108
DB 1111 LAHWSPAKLLQWSSATAYGSTVSKRVAWHYDEEKIEFEMNTGTNVDTKGMTSNFPVDL 1170
QY 109 -----WF----- 110
DB 1171 SDYPKSLMYANRLDHRVPETDMTRFHVGSKLIIVAMSSWLQKASGLPYTOTLQDHLNS 1230
QY 111 -----DFYSTQ-----LR 118
DB 1231 LKEFNLQNGMLPDFHLPENLFLKSDGRVKYTLNKNLSKIEIPLPGGKSRDLKMLETVR 1290
QY 119 KPARY-----VYSQ-YNHTA----- 132
DB 1291 TPALHFKSVGPHLPSREFQVPTFTIPKLYQLQVLLGLDLSTNVYSNLNMSASYSNGN 1350
QY 133 -----KTITF----- 137
DB 1351 TSTDHPSLRARYHMKADSVVDLLSYNVQSGETTYDHRKNTFTLSCDGSURHKFLDSNIKF 1410
QY 138 ----- 137
DB 1411 SHVEKLGNNPVSKGLLIFDASSSWGPMQASVHLDSKKQHLFVKEVKIDGQFRVSSFVA 1470
QY 138 -----RPPPCGRVP-----SMTCLSEM----- 154
DB 1471 KGTYGLSCQDPNTGRLNGESNLNRSYLOQNTQITGRYEDGTLSLTSTSLQSGIKN 1530
QY 155 -----LNVSCKN----- 161
DB 1531 TASLKYENVELTKSDTNGKYKNFATSNQMDMTFSQNALLRSEYQADYESLRFFSLSG 1590
QY 162 -----DTGE 165
DB 1591 SLNSHGLELNADILGTDKINSGAHKATLRIGODGISTATSATTNLKCSLLVLENELNAELGL 1650
QY 166 QCG-----NF----- 171
DB 1651 SGASKMLTTNGRFRHNAKFLDGAALTELSGAYQAMILGVDSKNIENFKVSOEGLK 1710
QY 172 -----TTF 174
DB 1711 LSNMWSYAEMKFDHTNSLNTAGLSLDFSKLDNIYSSDKPYKQTVNLQLOPYSLVTTL 1770
QY 175 N-----PM----- 177
DB 1771 NSDLKYNALDLTNGKLRLEPLKLVAGNLKAYQNNKIKHIVAISSAALSASYKADTVA 1830
QY 178 -----FNVPR----- 183
DB 1831 KVOGVESHRLNTDIAGLASAIDMTNYSNLSLHFSNVFRSVMAPFTMTIDAHTNGKL 1890
QY 184 ----- 183
DB 1891 ALWGEHTQOLYSKFLKAEPLAFTPSHDYKGSTSHHLVSRKSISSAALEHKVSALLTPAQ 1950
QY 184 -----WNTKLYVG-----PTKV----- 195
DB 1951 TGTWKLKTQFNNEYSQDLDAINTKDKIGVELTGRTLADTLTLDSPIKVPLLLSEPINII 2010
QY 196 -----NVDSTTI-----YF----- 204
DB 2011 DALEMDRAVEKPEFTIIVAFVKYKQDVHSINLPFFETLQEVFERNROTIIWVENVOR 2070
QY 205 -----LG-----LTALLRYA----- 215
```

Db 2071 NLKHINIDQVRKYRAALGKLPOQANDYLNSFNWQVSHAKEKLTALTKKYRITENDIQ 2130
Qy 216 ----- 215
Db 2131 IALDDAKINFNEKLSQLQTTMIQPDQYKOSYDLHLDKIAIANIIDBIIEKJLSLDEHYH 2190
Qy 216 ----- 215
Db 2191 IRVNLVKTIHDLHLFIENIDFNKSGSSTASWIQNVDTKYQIRIQIEKQLQQLKXHQINID 2250
Qy 216 -----QRNCTHSFVLNMSRNL-----FRVPKYING- 242
Db 2251 IOHLAGLKGHIEAIDVRVLLDQLGTTISFERINDVLEHVKHVFVINLIGDFEVAEKINAF 2310
Qy 243 -----TKLKNVWRKLKR--KQAPVKEQEFK----- 265
Db 2311 RAKVHELIEREYVDQIQVLMKDLVELTHQYKLETKIQKLSNVLQQVKIKINDYPEKLVGFI 2370
Qy 266 ----- 265
Db 2371 DDAVKLNLSEKFTFIEDVNKFLDMLIKLSPDYHQFVDETNDKIREVTQRLNGEIQAL 2430
Qy 266 -----KAK----- 268
Db 2431 ELPQKABALKFLEETKATVAVYLESQDKTITLIINWLOEALSSASLAHMKAKPRETLE 2490
Qy 269 -----KTOSTTTTPYS-YTTSAAANVTNV-TYSITTAAARV 303
Db 2491 DTRDMYDMIOQELQYLSLQGVSTLVYISDWMTLAKNLDTFABEQYSIQDWAKRM 2550
Qy 304 STSTIAVRPDSF---MKSIMATQ----- 324
Db 2551 KALV-----BOGTVPEIKTILGTMPAFEVSLQALQKATQTPDFIVPLTDLRIPSVQIN 2605
Qy 325 ----- 324
Db 2606 FKDLKNIKIPSRFSTPEFTILNTHIPSTFIDFVEMKVKIIRTIDQMONSELQMPVPDIY 2665
Qy 325 LRDL----- 328
Db 2666 LRDLKVEDIPLARITLPDFELPEIAPEFIPTLNLNDFQVPLHIPEQLPHISHTIEV 2725
Qy 329 ----- 328
Db 2726 PTFGLYSILKIQSLPFLTDANADIGNGTTSANEAGIAASITAKGSKLEVLNDFDQANA 2785
Qy 329 ----- 328
Db 2786 QLSNPKINPLAKESVKFSKYLRTEHGSMLPFGNAIEGKSNVVASLHTEKNTLELSNG 2845
Qy 329 -----ATWVYT 334
Db 2846 VIVKINNQLTDSNTKYFHKLNIPKLPDSSQADLRNEIKTLLKAGHIAWTSSGKGSWKA 2905
Qy 335 T-----LRYRQN----- 341
Db 2906 CPRFSDGTHESQISFTIEGPLTSFGLSNKINSKHLRVNQNLVYESGSLNFSKLEIQSOV 2965
Qy 342 -----PF-CBPSRNR 350
Db 2966 DSQHVGHSVLITAKGMALFEGGKABFTGRHDAHLNGKVI GTLKNLSLFFSAQPFETASTNN 3025
Qy 351 -----TAVSEFMKN----- 359
Db 3026 EGNLKVRFPLRLTGKIDFLNNVALPLSPAQASQVARSARFNQYKKNQNFSAENNENIME 3085
Qy 360 THVLIRNET-----PYTYGT----- 375
Db 3086 AHVGEINGEANLDFLNIPLATIPEMRLPYTIITPPLKDFSLWEKTLGKLEPLKTKQSFDLS 3145
Qy 376 -----LDMSSLYNET----- 386
Db 3146 VRAQYKKNKXHSITNPLAVLCEFTSQISKSFDRHFEKRNALDPVTKSYNETIKIPDK 3205

Qy 387 ----- 386
Db 3206 YKAEKSHDELPRTFQIPGYTVPVNVVESPFTIEMSAFGYVPPKAVMSPSFILGSDVRV 3265
Qy 387 ----- 386
Db 3266 PSYTLILPSLELPVLHVPRNLKLSLPHPKELCTISHIFIPAMGNITYDFSKSSVITLNT 3325
Qy 387 -----MFVE---NK 392
Db 3326 NAELFNQSDIVAHLLSSSSVIDALQYKLEGTTLTRKRGKLATALSLSNKFVEGSHNS 3385
Qy 393 TAS-----DSNKTTP-TSPSGFORFT----- 413
Db 3386 TVSLTTKNNEVSAKTTKABIPILRMNFQOELNGNTKSKPTVSSSMFEKFDNSSLMYST 3445
Qy 414 ----- 413
Db 3446 AKGAVDHKLSLESLTSYFSIESSTKGVKSVLSREYSGTIASEANTYLSKSTRSSVKL 3505
Qy 414 -----ID----- 415
Db 3506 QGTSKIDDIWNLEVKENFAGEATLQRIYSLWEHSTKNHLQLEGLFTNGEHTSKATLELS 3565
Qy 416 ----- 415
Db 3566 PWOMGALQVHASQSPSFHDPDLGOEVALNANTKNQKIRWKNVRIIHSGSFQSQVLSN 3625
Qy 416 -----PLWDYL----- 421
Db 3626 DQEKALHDIAGSLEGHLEHLKNIILPVVDKSLMDFLKDVTTSIGRRQHLRVSTAFVTK 3685
Qy 422 ----- 421
Db 3686 NPNGYFSIPVKVLADKFTTGLKLNLSVLVMPFTFVPTDLQVPSCKLDFREIQYK 3745
Qy 422 -----DSLLELDEI----- 430
Db 3746 KLRTSPALNPLTLPVKFPEVDVLTYSQPEDSLIPFETITVPESQLTVSQFTLPKVS 3805
Qy 431 ----- 430
Db 3806 DGIAAALDNVANKIADPELPTIIVPEQTIETIPSIKFSVPAGIVIPSFQALTARFEVDSP 3865
Qy 431 -----R 431
Db 3866 VYNATWSASLKNKADYVETVLDSTCSSTVQFLEYELNVLGTBKIBDGTLAGTKTGLAHR 3925
Qy 432 NPS-----LRSPT----- 439
Db 3926 DSAEYBEDGKEGLEWEGKAHLNIKSPAFTDLHLRYQKDKKGISSAASPAVGTGMD 3985
Qy 440 ----- 439
Db 3986 MDEDDFSKWNFYSPQSSPDKLTI FKTELRVRESDEBTQIKVNWEESAAGLTLTKD 4045
Qy 440 -----YVN----- 442
Db 4046 NVPKATGVLYDVNKYHWEHTGLTLREVSKLRRNLQNNAEVYQGAIRQIRDDIDVRFOK 4105
Qy 443 ----- 442
Db 4106 AASGTTGYQEWKDKAQLYQELLTOEQOASFOGLKQNVFDGLVRVTQKFHKVKHLIDS 4165
Qy 443 ----- 442
Db 4166 LIDFLNFRFOGPGGIYTTREELCTMFIREVGTVLSQVSKVHNGSEILFSYFODLVIT 4225
Qy 443 ----- 442
Db 4226 LPFELRKHLIDVISMVRELLKDLSKAEQVFKAIOSLKTTEVLRLNLODLOFTQFOLIED 4285

QY 443 ----- 442
Db 4286 NIKOLKEMKFTYLINYIQDEINTFNDYIPVFKLLKENLCLNLRKFNFIQNELOEASQ 4345
QY 443 ----- 442
Db 4346 ELQIQHYMALREYFDPISVGTWKYYELEEKIVSLIKNLVALKDPHSEYIVSASNP 4405
QY 443 -----LTPP----- 446
Db 4406 TSQSSQVEQFLHRNIQELSLTDPDGKKEIAELSAFAQBIKSAIATKTKIISDYH 4465
QY 447 -----EHRVNLSTN 458
Db 4466 QPFRYKLQDPSDQSDYYEKFAIESKRLIDLSQN 4500

RESULT 9
US-08-750-717-2
; Sequence 2, Application US/08750717
; Patent No. 6180109
; GENERAL INFORMATION:
; APPLICANT: MOORMANN, Robertus J. M.
; APPLICANT: VAN RIJN, Petrus A.
; TITLE OF INVENTION: Nucleotide Sequences of Pestivirus
; TITLE OF INVENTION: Strains, Polypeptides Encoded by These Sequences and Use
; TITLE OF INVENTION: Thereof for Diagnosis and Prevention of Pestivirus
; TITLE OF INVENTION: Infections
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750,717
; FILING DATE: 24-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94201743.5
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/NL95/00214
; FILING DATE: 16-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: BO 39123
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-750-717-2

Query Match 17.4%; Score 426; DB 3; Length 3898;
Best Local Similarity 5.0%; Pred. No. 0.0059;
Matches 191; Conservative 88; Mismatches 171; Indels 3373; Gaps 61;
QY 1 MGRKEMM-----VRDVPK----- 13
|||:|:

Db 18 MGVEEPVYDTCRPLPGDPSEVHPQSTLKLPHDRGRGNIKTTLKNLPRKDCRSGNHLGP 77
QY 14 -----MFVLISI 20
Db 78 VSGIYVKPGPVFYQDYMGPVYHRAPLEFDEQFCEVTKRIGRVTSKGLYHYVICDG 137
QY 21 SPILLV-----SPINC----- 30
Db 138 CILLKLAKGEPRTLKWIENFTDCLPWVTS CSDGASGSKKPKDRINKGKLIAPKEHE 197
QY 31 -----KMSKALY----- 38
Db 198 KDSRTRPPDATIVVEGVKYQVKKKVGKNTQDGLYHNKPKPSRKKLEKALLAWAVI 257
QY 39 ----- 38
Db 258 AIMLYQPVEAENITOWNLSDNGTNGIOHAMYLGRVNSLHGIWPGBEICKGVPTHLATDVE 317
QY 39 ----- 38
Db 318 LIKEIQGMDASEGTNYTCCKLQRHEWNKHGWCNWHNIDPMIOLMNRTOADLAEGPPVKEC 377
QY 39 -----NRP----- 41
Db 378 AVTCRYDKDADINNVTOQRNRPPTLTGCKKGNFSPAGTVIESPCNFVSVEDTLYGDEH 437
QY 42 -----WRG----- 44
Db 438 CGSLQLQDAALYLVDGWTNTIENARQGAARVTSWLGRQLTAGKRLGRSKRTWFGAVALSP 497
QY 45 -----L 45
Db 498 YCNVTSKIGYIWTNNCTPACLPKNTKIIGPKGFTDAEDGKILHEMGHLSBFLLSLV 557
QY 46 VLS-----KICKY-- 53
Db 558 VLSDFAPETASALYLILHVVIPOPHDEPEGCDTNQLNLTVELRTEDVIPSSVWNVGKYVC 617
QY 54 -----KLDQLKLEILRQLE----- 67
Db 618 VRPDWMPYETEVALLFEEVGVVGLALRALRDLTRVWNSASTIAFLICLIKVLRGQIVQG 677
QY 68 -----TTISTKYN----- 75
Db 678 VVWLLVVTGAOGLACKEDYRVAISDTEIGLIGAGGLTTTWKEYNHDQLQNDGTVKASC 737
QY 76 -----VSKQPVKNL----- 84
Db 738 VAGSFVKTALNVSRKYLASLHKKALPTSVTFELLEPDGTPNPSSTEMGDDFRSGLCFPTS 797
QY 85 -----TWNTE----- 89
Db 798 PVVKGKYNITLLNGSAFYLVCPIGWTVIETAVSPTTLRTVEVTKFRDKPPPHRMDCV 857
QY 90 -----FPQYVI----- 95
Db 858 TTTVENEDLYCKLAGNWTCKVGEPPVYTGVLVKQRCWCGPDPGDPGLFHYPIGRKILA 917
QY 96 -----LAGP 99
Db 918 NETGRIVDSTDCNRDGVVISTEGSHECLIGNTVTKVHASDERLGMPCRPKEIVSSAGP 977
QY 100 IQNYSITY-----LWFPDYSTQLRKP----- 120
Db 978 VKKTSCTFNKTKLNRYVEPRDSYFQYMLKGEYQWFDLDTDRHSYPAEFVVLVVV 1037
QY 121 -----AKYV----- 124
Db 1038 ALLGGRYVLWLIYTVVLTQLAAGLPQGEVVLIGNLITHTDIEVVVYVFLLLYLVMRD 1097
QY 125 -----YSQVNHAKTIT----- 136
Db 1098 EPIKKWILLFHAMTNNPVKTTITVALLMVSVAKGKIDGGWQRLPGTGFDIQLATVIV 1157
|||:|:

QY 137 ----- 136
Db 1158 VAVMLAKRDPPTVPLVITVAPLRTAKMTNGLSTDIAIAIVSAALLTWIISDYRYKTV 1217
QY 137 -----FRP----- 139
Db 1218 LOYLIVTGTGIFLIRVLKGIGELDLHTPLPSHRPLFFILVYLSTAVVTRWNLDIAGLL 1277
QY 140 ----- 139
Db 1278 LQCVPTLLMVTWADILTLILPTVELTKLYLKEVRIGAEKGMWKTNFRVNDIYE 1337
QY 140 ----- 139
Db 1338 VDQAGEGYLPPSKQKTSMTGTMPLLIKALLISCVSNKQFIYLLYLIFEVSYYLHKI 1397
QY 140 ----- 139
Db 1398 IDEIAGGTNFIISRLVAAIEVNAFONEEVRLKGLFLLSRVKELIHKHVRNEVMVRW 1457
QY 140 -----PPCGRV-PSMTC----- 150
Db 1458 FGDEEYVCMPEKLVGLVKAATLSKNKHCICTVCEDREWGETCPKCGRFGPPMTCGMTLA 1517
QY 151 ----- 150
Db 1518 DFEKHVKRIFFREDOSEGPVREEYAGYQYRARGQLFLRNLPLVATKVKQMLLVGNLGE 1577
QY 151 -----LSEMLNV 157
Db 1578 VGDLEHLGWLVRGPAVCKVTEHEKCTTSMMDKLTAFPGVMRGTTPRAPRPPTSLIKI 1637
QY 158 SKRNDTG-----EOG 167
Db 1638 RRLEGTWAYTHOGGISVDHVTGCKDLLVCDTWGRTRVVCQNNKMTDESEYGVKTDG 1697
QY 168 C-----GNFT-----TFNPMFF-----NVPRWN----- 185
Db 1698 CPEGARCVFNPPEAVNISGTGAMVHLQKTGGEFTCVTASGTAFDFDLKILKMGSLPIF 1757
QY 186 -----TKLYVG-----PTKVNDVDSQTI----- 202
Db 1758 EASSGRVVRGVKGNEDSKPTKLMGSIQTVSKSTTDLTEMVKKITITNGRGEFRQITLAT 1817
QY 203 ----- 202
Db 1818 GAKTTELPRSVIEEIGRHKRVLVLIPLRAAESVYQYMRQKHPSTAFNLRIGEMKEGDM 1877
QY 203 -----YFL-----GLTALLRYA-----QRNCT-----HSFVLNAMSR 231
Db 1878 ATGITVASYGYFCMPQPKLRAAMVEYSFIFLDEYHCATPEQLAIMGKIHRF-----SE 1931
QY 232 NL-----FRVPKYINGTKLN----- 247
Db 1932 NLRVAVMATPVGTVTGQKHPIEBFIADPVMMKGDGLGSEYLDIAGLAKIPVEEMKSNML 1991
QY 248 -----TMRKLKRX-----QAP----- 258
Db 1992 VFVPTERNMAVETAKKLKAGYNSGYYSGEDPSNLRVVTSQSPYVVVATNAIESGVTLPD 2051
QY 259 -----VKEQFEKKAK----- 268
Db 2052 LDVVVDGLKCEKRIIRLSPMPPIFVTLGRMAVTTIGEQAQRGRVGRVKPGRYRSQETP 2111
QY 269 ----- 268
Db 2112 VGSKDHYDLLQORYIEDGINITKSPREMNVDSLYEDSLMITQLBILNNLLISDEL 2171
QY 269 -----KTQSTTTTPYFSYT----- 281
Db 2172 PMAVKNMARTDHPPEIQLAYNSYETQVPLFPKIKNGEVTDSYDNYITPLNARKLGDDVP 2231

QY 282 ----- 281
Db 2232 PYYATEDEDLAVELLGLDWPDPGNOQTIVETGRALKQVVGSLSTAENALLVALLFGYVGYOA 2291
QY 282 ----- 281
Db 2292 LSKRHIPVVTDIYSIEDHRLDTHLQYAPNAIKTEGKETELKELAQGDVQRCVBEAMTVY 2351
QY 282 -----TSAALNVTTNTY----- 294
Db 2352 AREGIOFMKSQALKVKKETPTYKETMDTVTDYVVKFMEALADSKEDIKYGLMGTHTALYK 2411
QY 295 -----SITTAARRVSTTIAY-----RPD-----S 314
Db 2412 SISARLGGETAFATLVVWKLAFGGSIAHDHVKQAATDLVVYIINRPQPGDTETQODGR 2471
QY 315 SFWKSIMATQLRDLATVWYTLRYR----- 339
Db 2472 KEVASILLASA---LATYTKSNWYNNLSKIVEPALATLPYAATALKLPAPTRLESVVILS 2528
QY 340 -----QNPP-----CEPSRNRATVSE--- 355
Db 2529 TAIYKTYISIRKGSDELGLGTGVSAMEIMSQNPVSVGVIAVMLGVGCAVAAHNAIEASEOK 2588
QY 356 -----FMKN-----THVLRNETP-----YTIYGTL----- 376
Db 2589 RLLMKVFVNFLDQAATDELVK-ESPEKI INALFEAVQTVGNPLRLVYHLGVFYKGWE 2647
QY 377 -----DMSSLYYNETMFMVENKTASDNKT--- 400
Db 2648 AKELAQRTAGRNLFITLIMFEAVELLGVDSEKVRQLSSNYILELY---KFRDSIKSSV 2703
QY 401 -----TPTSPSMGFQR----- 411
Db 2704 REMAISWAPAPSCDWTPTDDRIGLUPDNHFQVETKPCPGYKMKAVKNCAGBELRLEBEG 2763
QY 412 ----- 411
Db 2764 SFLCRNKEGRGSRNRYVKTYDDNLEIKPVIRMEGHVELYKYGATIKLDFNNSKITLAT 2823
QY 412 ----- 411
Db 2824 DKWEVDHSTLVRVKRHTGAGYHGYLGEKPNHKLIERDCATITKDKVCFMKMRGCAF 2883
QY 412 -----TFI----- 414
Db 2884 TYDLSLHNLTRLIELVHKNNLEDKEIPAATVTTWLAYTFVNEIDIGTIKPAFGEKVTLEMQ 2943
QY 415 ----- 414
Db 2944 EBITLQPAVVVDTTDVAVTWVGEAPTMTTGTPTVFTSSGSLKQQVQLKLGVBEGYPG 3003
QY 415 -----DPL----- 417
Db 3004 TNPORASLHEAIOGADERPSVLILGSDKATSNRVKTAKNVKYVGRDPLEVRDMMRGKI 3063
QY 418 -----WDYLDLSLLFLDEIRNFSLSRPTVYVNL- 444
Db 3064 LVVALSRVDNALLKEVDYKGTFLTREALEALSGLRPPKKNTIKABAQWLLCPEDQMBELP 3123
QY 445 -----PP----- 444
Db 3124 DWFAAGEPIFLEANIKHIDRYHLVGDIAITIKAKQLGATDSTKISKEVGAKVYSMKLSNW 3183
QY 445 -----PP----- 446
Db 3184 VMQENKOGNLTPLFEELLQQCPGQGNKTAHMVSAYOLAQGNWMTSCHVFMGTVSARR 3243
QY 447 -----EHR----- 449
Db 3244 TKTHPYEAVVKLREIVEEHKMKTLCPGSSLRGHNDWIIIGIKIYQGNLRTKHLNLPKVAE 3303
QY 450 ----- 449

Db 3304 QLCREGRHNNVNTISSVNTATGIRLEKLPVVRAQTDPNTFHOAIRDKIDKEENLQTPG 3363
QY 450 -----RAVN----- 453
Db 3364 LHKLMVFNALKRPELESSYDAVEBELERGINRKGAGFPFRKNIGILDSKKNKVEE 3423
QY 454 ----- 453
Db 3424 IIDNLKGRNIKYETAIKNEKRDVNDWDTAGDFVDEKKPRVIOYPEAKRLAITKVMY 3483
QY 454 ----- 453
Db 3484 KWKQKPVVPGYEGTKPLFIQIDFKVKKEWQONPVAVSFDTKAWDTQVTTKOLELIRD 3543
QY 454 -----LSTNSNL----- 460
Db 3544 IQKYFKKKWHKFIDLTTHMSEVPVISADGEVYIRKQGGSGQPDTSAGNSMLNVLTVW 3603
QY 461 ----- 460
Db 3604 YAFCEATGVPKSDFRVAKIHVCGDDGFLITERALGEKFAKGVQILYBAGKPQKITEGD 3663
QY 461 ----- 460
Db 3664 KMKVAYQDDIEFCSHTPIQVRWSDNTSSYMPGRNTTILAKWATRLDSSGERGTIAYEK 3723
QY 461 -----W----- 461
Db 3724 AVAFSLLMYSWNPILRRICLLVLSTELQVKPKSTYYYEGDPSAYKEVIGHNLFDLK 3783
QY 462 -----W-----WL 464
Db 3784 RTSFEKLAKINLSMSVLGAWTRHTSKRLQDCVNIGVKGSGNL 3826

RESULT 10

US-08-222-617A-2
; Sequence 2, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-222-617A-2
Query Match 17.2%; Score 422; DB 2; Length 3778;
Best Local Similarity 5.2%; Pred. No. 0.0074;
Matches 193; Conservative 69; Mismatches 171; Indels 3271; Gaps 55;
QY 8 VRDVK-----MFVLISISF----- 22
Db 76 VNQVPERCDLSGLTDDSTRYQLASTGFGDASAAQERLMTVPVDVHAALQELCLERRSV 135
QY 23 ----- 22
Db 136 GSVINFSVHQLKGFNGTHTTITASLHREQLNQNSPSPVSVPTIVTHENRDGWSVAQAV 195
QY 23 -----LLVSFIN-----CKV- 32
Db 196 ESIEAARGSEKESVTAIDSASSLVKMLFDLLVSVFVDDARIPCFDFPLAVIVRECDAN 255
QY 33 -----MSKALYN-----RWRG-- 44
Db 256 LSLTRFSDCLFNEETICNFTDALMILLAEAVIGRVPVADIELLSAEKQOLEEWNNTD 315
QY 45 ----- 44
Db 316 GEYPSKRLHLIEEVRHEDKIAVVCDERELTYGELNAQNSLARYLRSIGILPEQLV 375
QY 45 -----LVLSKGKYK-----LDQLK----- 59
Db 376 ALFLDKSEKLVITILGVNWKSGAAVYPIDTPYDPRVRFVLDOTKARAIASNQHVRLQR 435
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TMNTEFQ----- 92
Db 436 EVIGDRNLICINLEPLLASAODSSKFFAHNLDLPLTSQQLAYVYTSGTGTFPGKIPK 495
QY 93 -----YVILAGP----- 99
Db 496 QHTNVVNSITDLSARYGVAGQHHEAILLFSACVPFVFRQTLMAVNGHLLAVINDVEKY 555
QY 100 -----IQNYSITYL-----WDFP----- 112
Db 556 DADTLPPFIRHSITYLNGTASVLQEDFSDCPSLNRIILVGENLTEARYLALRQRFKNR 615
QY 113 -----YSTQLRPKAK-----YVYSQYNHTAKTITFRPPPCG 143
Db 616 ILNEYGFTESAFVTKIFDPESTRKDTSLGRPVNRVNCYILN-----PSLK 662
QY 144 RVPMTCLSEM-----LNYSK----- 159
Db 663 RVP-IGATGELHIGGLGIGSKGYLNPRLTPHPIPNFPQDCEKQGLGINSLMYKTDGLAR 721
QY 160 -----RNDTGEQCGNFTFNPMPFNPRWNTKL----- 188
Db 722 WLPNGEVEYLGRADFQIKLRIGIRBPGE-----IETMLAMYPRVTSLVVSKKLNRG 773
QY 189 -----YVGPTK-----VNYD----- 198
Db 774 PEETTNEHLVGYVCDSSASVSADLLSFLEKLPYIMPTRLVQLSQIPVNVNGKADLRA 833
QY 199 -----SOTVFLG----- 206
Db 834 LPVDISNSTEVRSLRGDTETALGEIWDVLGARQSVSRNDNFRGLGHSITICLIQILA 893
QY 207 -----LTALLRYAQRNC-----THSFYL 225
Db 894 RIRQRQLSVSISVEDVFATRTLERWADLLQNKQEKCDKPHEAPTELLENAATDNIYL 953
QY 226 VNAMSR----- 231

Db	954	ANSLQOQGVHYHLSKMEQSDAYVMQSVLRYNTTSLPDLFORAMKHAQQSPALRLRFSWE	1013
Qy	232	-----NLFV-----	236
Db	1014	KEYFOLLDDQPPDLWRFLYFTDVAAGAVEDRKLEDLRRQDLTERPKLDVGRLFVYLIKH	1073
Qy	237	-----	236
Db	1074	SENRFCLFSCHHAILDWSLPLLEKVVHETYLQLLHGDNLTSMDPPYTRTQYLAHR	1133
Qy	237	-----PKY-----	239
Db	1134	EDHLDFWAGVVQKINERCDMALLNERSRYKVLADYDQVEQRHVHTIALSGDAWLADLR	1193
Qy	240	-----INGT-----	243
Db	1194	QTCSAQGITLHSLIQFVWHAVLHAYGGTHITIGTTISGRNLPILGIERAVGPYINTLPL	1253
Qy	244	-----KLNWTKLKRKAQVK-----	260
Db	1254	VLDHSTPKDKTIMEATEDVQAKVMVMSRGNVELGRHLKTDLKHGLFDSLVLNYPNLD	1313
Qy	261	-----	260
Db	1314	KSRTLEHQTELGSYIEGGTEKLNPLAVIAREVETGGFTVSYCYASELPEEVMISELLH	1373
Qy	261	-----EOPEKKAKK-----	269
Db	1374	MVQDTLMQVARGLNEPVGSLEYLSSIQLEQAAWNATEAFPPDTTILHEMFENEASQPKD	1433
Qy	270	-----	269
Db	1434	IADVBEETSITYRELNERANMAHQLRSDVSPNPNVIALVMDKSEHMTIVILAVKSGG	1493
Qy	270	-----	269
Db	1494	AYVPIDPGPNDRIOVILEDTQALAVIADSCYLPRIKGMASGTLIYSPVLPANPDKWS	1553
Qy	270	-----TOSTTTPFSYT-----	281
Db	1554	VSNPSPLSRSTDLAYIYTSGTGRPKGVVEHGVNVLQVLSKVPGLRDTDDDEVILSF	1613
Qy	282	-----TSAALNVT-----	294
Db	1614	SNYVDFHVEQMTDALINGQTLLVLNDGMRGDKERLYRIEKNRVTYLSGTFSVSMYEF	1673
Qy	295	-----SITT-----	298
Db	1674	SREFKDLRRVDCVGEAFSEPVDPKIRETFHGLVINGYGTVEVSITTHKRLYPFERMDK	1733
Qy	299	-----	298
Db	1734	SIGQQVHNSTSYVLNEDMKRTPIGAVGELYLGEGVVGVGNRADVTAERFIPNPQSEE	1793
Qy	299	-----	298
Db	1794	DKREGNSRLYKTDLVRWIPGSGSEVBYLGRNDFQVKIRGLRIEVEIEAILSSYHGIK	1853
Qy	299	-----AARR-----	302
Db	1854	QSVVIACDCEGAQKFLGVYVADAALPSAAIRRFMQSRLPGVMVPSRLIIVSKFPVTPS	1913
Qy	303	-----	302
Db	1914	GKLDTKALPABEESIDVVPFRSEIERSLCDIWAELLEMBHEEIGIYSDPFSIGDLSLK	1973
Qy	303	-----VSTSTI-----	308
Db	1974	STKLSMIIHESFNRAVSALFCHRTVEAQTHLIILNDAADVHEITPIDCNDTQMIPVSR	2033
Qy	309	-----AYRPDSGF-----	316
Db	2034	QERLLFIHBFENGSNAYINDAAFELPGSVDSILLEGALRGNLARHEALRTLVLKDHATGI	2093
Qy	317	-----	316
Db	2094	YLOKVLSPDEAQGMFSVNVDTAKQVERLDQETASLSQHVFRLLDDELPEARILKLESGGL	2153
Qy	317	-----	316
Db	2154	YLILAFPHHTCFDAWSLKVFEQELRALYAAALQTKSAANLPALKQAQYKEYALVHRRQLSGD	2213
Qy	317	-----	316
Db	2214	RMNRLSDFWLRLIGLEPLQLITDRPRPVQFYKDGDDLSIELSKKETENLGRVAKRCKSS	2273
Qy	317	-----	316
Db	2274	LYVVLVSVCVMLASYANOSDVSGIPVSHRTHPOFOSVIGFVNVLVLRVDISQSAICG	2333
Qy	317	-----MKSIMATQLRD-----	327
Db	2334	LIRRVMKELVDAQLHQDMPFQBVTKLQVDNDPSRHPVQVNFNFRANGEDHARSEDE	2393
Qy	328	-----	327
Db	2394	GSLAFNQVRPVQVDSVAKFDLNLATVTELESGLRVNFNYATSLFNKSTIQGFLLHYELL	2453
Qy	328	-----	327
Db	2454	ROLSLSAEGINEDTQLSLVRPTENGDLHLPLAQSPATTAEEOKVASLNQAFEREAFLA	2513
Qy	328	-----LATW-----	331
Db	2514	AEKIAVQGDRLASYADLNGQANQLARYTQSVSCIGADDGIALMLEKSIDTICILAIWK	2573
Qy	332	-----	331
Db	2574	AGAAVPLDPTYPGRVOLILEEIKAKAVLVHSSHASKCERHAKVIAVDSPAIETAVSQ	2633
Qy	332	-----	331
Db	2634	QSAADPTIASILGNLAYIIFTSGTSKPKGVLVQKAVALLRDLALRERYFGDRDCTKHGV	2693
Qy	332	-----	331
Db	2694	LFLSNYVDFSVQVLVLSVLSGHKLIIVPPAEFVADDEBFYRMASHTGLSYLSGTPSLLOKI	2753
Qy	332	-----VY-----	333
Db	2754	DLARLDHLOVTAAGEELHATQEKORRRFNGPIYNAVGVTTVTYVNIIAEFTTNSIFEN	2813
Qy	334	-----TTLRYRQNPFCBP-----	346
Db	2814	ALREVLPGTRAVVLTAALQPVFPDVGELYLAGDVTVTRGYLNQPLLTDOORFIPNPFCKEE	2873
Qy	347	-----SR-----NRTAVSE-----	355
Db	2874	DIAMGRFARLYKTGDLVRSFRNQOQPOLEYLGRGLDLOIKMRGVRIEISEVQNVLTSPPG	2933
Qy	356	-----FMKNTHVLI-----RNET-----	368
Db	2934	VREGNAVAKYENDTYSTRHSLVGYTTDTNETVSEADILTFMKARLTYVWPVSHLCLE	2993
Qy	369	-----PVTIYGLDMSSL-----	381
Db	2994	GALPVTINGKLDVRLPEIINDSAQSSYSPPNRIIEAKMCRLESALGMERCGIDDDLFK	3053
Qy	382	-----	381
Db	3054	LGDSITSLHLVAQIHNOVGCKITVRDIFEHRTARALHDHVMKDSRNSVNTQFRTEQGP	3113
Qy	382	-----YNETMFV-----	389
Db	3114	VIGEAPLLPIQDWFLSKALQHPMTWNHTFYVTRPELVDVLSAAVRDLOQYHDFVRLK	3173

390 ----- 389
3174 REEVGVQSAEDFSPAQLRVNLVKDVGSAANNEILDGMQGFLENPGISGYLHG 3233
390 ----- ENKTASDSN 398
3234 EDRSARVWFVHHMAIDTVSQIILVRLQTLRYNRGSLGSGSSPRQWABAIQNTKADSE 3293
399 KT----- 400
3294 RNHWKLVMTASSISALPTSTGSRVLSRSLSPKTSALIQGGIDRQDVSVYDSLTSV 3353
401 -----TPTSPSM-----GQRTFID----- 415
3354 GLAQHIAPTGPSMTIEGHGREVDQTLDSRTMGWFTTMYPPFIPRLSTENIVQGWA 3413
416 ----- 415
3414 VSRPQVARGVGYGLYGTQHPQLQVTVNYLGQARKQSKPEWLVAVGONEPEYGL 3473
416 ----- 415
3474 MTPEDKDRSSAVDTAVCIDGTMIIIDVDSAWSLESEQFISSIEGLNKILDGRASQ 3533
416 -----PLWDYLD----- 422
3534 TSRPPDVQPAETVTPFYEYLEPPRQGTFLPLPPGEGGAESYFNNIVKRLQTNMVFN 3593
423 -----SLLFDEIRNFSLRSP----- 438
3594 NYLHSLRLRTFELAEMLDQVRGIQPHGPHYHFGISFGGILAMEMRSLVADSEKIGF 3653
439 -----TYVN-----LTPPEH----- 448
3654 LGIIDTFVNRGATRTIGLGTLDPIHHIYNPDPAFQRLPSATDRIVLFKMRPNK 3713
449 -----RR-----AVNLSTSNLW 462
3714 YESENQRRLYEYDRTLRGLDLSLLPSDSVQLVPLTDTHFSW 3757

RESULT 11

US-08-222-617A-12
; Sequence 12, Application US/08222617A
; Patent No. 5883879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan F.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palissa, Harriet
; APPLICANT: Van Liempt, Henk
; TITLE OF INVENTION: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222, 617A

FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
REFERENCE/DOCKET NUMBER: 97,157
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3666 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Penicillium chrysogenum
FEATURE:
NAME/KEY: Protein
LOCATION: 1..3666
OTHER INFORMATION: /label= region
OTHER INFORMATION: /note= "Region of ACV Synthetase; aa 62-372"
US-08-222-617A-12

Query Match 17.1%; Score 420; DB 2; Length 3666;
Best Local Similarity 5.2%; Pred. No. 0.008;
Matches 188; Conservative 68; Mismatches 163; Indels 3211; Gaps 53;

QY 8 VRDVPK-----MFVLISISF----- 22
DB 15 VNQVPERCDLSGLTDTSTRYQLASTGFGDASAAQERLMTVPVDVHAALQELCLERRSV 74
QY 23 ----- 22
DB 75 GSVINFVHQMKGFGNGTHTITASLHREQLNQSSPSWVSPITVTHENRDGWSVAQAV 134
QY 23 -----LLVSPIN-----CKV- 32
DB 135 ESTEAAARGSEKSVTAIDSAASLVKMGFLFDLLVSFVDADARIPCFDFPLAVIVRECDAN 194
QY 33 -----MSKALYN-----RPMRG-- 44
DB 195 LSLTLRFSDCLFNEETICNFTDALNILLAEAVIGRVTPVADIELLSAEQKQLEWNTD 254
QY 45 -----LVLSKIGKYK-----LDQK- 44
DB 255 GEYPSKRLHLLJEVVERHEDIKAVVCDRELTGELNAGQNSLARYLSIGILPEQLV 314
QY 45 -----LVLSKIGKYK-----LDQK- 59
DB 315 ALFLDKSEKLIIVTLGVKWSGAAVPIIDPTYPDERVRFVLDLTKARAIASNOHVERLQR 374
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TWNTBPPQ--- 92
DB 375 EVIGDRNLCTIRLEPLLASLAQDSKFPANLDDPLTSQQLAYVTVTSGTTFPKGIFK 434
QY 93 -----YYILAGP----- 99
DB 435 QHTNVNVSITLSARYGVAQHAEIILFSACVFEPVROTLMALVNGHLLAVINDVEKY 494
QY 100 -----TONYSITYL-----WDFP----- 112
DB 495 DADTLFPFIRRHISITYLNGTASVLYQYDFSDCPSLNRIILVGNLNTAARYLALRQRPKNR 554
QY 113 -----YSTQLRKPAK-----YVYQYNHTAKTITFRPPPCG 143
DB 555 ILNEYGFTSAFTALKIFDPESTRKDTSLGRPVNVKCYLN-----PSLK 601
QY 144 RVPMTCLSEM-----LNVSK----- 159
DB 602 RVP-IGATGELHIGGLGISKGLNLRPELTPIHPFIPNPFQDCEKQLGINSIMYKTGDLAR 660
QY 160 -----RNDTGEQCCGNFTTFNPMFFNVPRWNTKL----- 188
DB 661 WLPNGEVEYLGRADFQIKLRGIRIEG-----IETMLAMYPRVRTLVSWSKLRNG 712

QY 189 -----YVGPTK-----VNVD----- 198
Db 713 PEETTNEHLGVYVCDASVSEADLLSFLKPLPRYMPITRLVQLSOIPVNVNGKADLRA 772
QY 199 -----SQTYYFLG----- 206
Db 773 LPAVDISNTEVRSLRGDTETALGEIWADEVILGARQRVSRRNDNFFRLGGHSITCIQLIA 832
QY 207 -----LTALLRYAQRNC-----THSFYL 225
Db 833 RIRQORLSVSISVEDVFATRTERMADLLONKQOQKCDKPHEAPTELEENAAATDNIYL 892
QY 226 VNAMSR----- 231
Db 893 ANSLQOGFVYHYLKSMEQSDAYMQSVLRYNTTILSPDLFORAKWHAQOSFPALRLRFSWE 952
QY 232 -----NLFRV----- 236
Db 953 KEVFQLDQDPLDMRFLYFTDVAAGAVEDRKLDRQDLTERFKLDVGRLFRVYLK 1012
QY 237 ----- 236
Db 1013 SENRFTCLFSCHAILDCWSLPLFEKVHETYQLLHGDNLTSSMDDPYRTRQYLAHR 1072
QY 237 -----PKY----- 239
Db 1073 EDHLDWAGVVQKINERCDMALLNERSRYKVQLADYDQEQORHVTTIALSGDAWLADLR 1132
QY 240 -----INGT----- 243
Db 1133 QTCSAQGITLHSILOFVWHAHLVHAYGGGTHITITGTSRNLPLGIERAVGPYINTLPL 1192
QY 244 -----KLKNTMRKLKRKQAPVK----- 260
Db 1193 VLDHSTFKDIMEAIEDVQAKVNMNSRGVNLGRHLKTDLKHGLFDSLFLVLENYPNLD 1252
QY 261 ----- 260
Db 1253 KSRTLEHQTELGYSGEGTEKLNPLAVIAREVETGGFTVSICYASELFEVWISSELLH 1312
QY 261 -----EQEKKAKK----- 269
Db 1313 MVQDTLMQVARGLNEPVGSLLEYLSSIOLEQLAAWNATEAFBPTTLHEMFENEASOKPDK 1372
QY 270 ----- 269
Db 1373 IAVVYEETSLTYRELNERANMAHQLRSDVSPNPNEVIALVMDKSEHMIWNILAVKWSGG 1432
QY 270 ----- 269
Db 1433 AYVPIDPGYPNDRIQYILETOALAVIADSCYLPRIKMAASGTLVPSVLPANPDSKWS 1492
QY 270 -----TQSTTTTPYFSYT----- 281
Db 1493 VSNPSPLSRSTDLAYIITYSGTTGRPGVTVHHGVNVLQVSLKVFGLRDTDDDEVILSF 1552
QY 282 -----TSAALNVTT-----NVTY----- 294
Db 1553 SNVVFDFHVBQMTDAIILNGOTLLVNDGMRDKERLYRYTEKRVITYLSTPSVWSMYEP 1612
QY 295 -----SITT----- 298
Db 1613 SRPKDHLRRVDCGEAFSEVPFKIRETFHGLVINGYGPTEVSITTHKRLYPPERRMDK 1672
QY 299 ----- 298
Db 1673 SIQQVHNSTSYVLNEDMKETPIGAVGELYLGEGVVRGYNRADVTAFRFPNPFQSEE 1732
QY 299 ----- 298
Db 1733 DKREGRNRLYKTGDLVRMIPGSSGEVEYLGRNDFQVKIRGLRIELGEIEAILSSVHGK 1792

QY 299 -----AARR----- 302
Db 1793 QSVVIAKCREGAQKFLVGYVADAALPSAIIRRFMQSRPLPGYVWFSRLILVSKFPVTPS 1852
QY 303 ----- 302
Db 1853 GKLDTKALPAEBESRIDVVPSEIERSLCDIWAELLEMHPPEIGIYSDFFSLGGDSLK 1912
QY 303 -----VSTSTI----- 308
Db 1913 STKLSFMIHESFNRAVSVALFCHRTVEAQTHLILDAADVHEITPIDCNDTQWIPVURA 1972
QY 309 -----AYRPDSF----- 316
Db 1973 QERLLFIHFENGSNAYNIDAFELPGSDVASLLEQALRGNLARHAEALRTLVLKDHATGI 2032
QY 317 ----- 316
Db 2033 YLQKVLSPDEAQMFSVNVDTAKQVERLQDEIASLSQHVFRLLDDELPEWARILKLESDDL 2092
QY 317 ----- 316
Db 2093 YLILAFHHTCFDAWSLKVFQEQLRALYAALOKTKSAANLPALKAQYKEYALYHRRQLSGD 2152
QY 317 ----- 316
Db 2153 RMENLSDFWLRLKLIIGLEPLQLITDRPRVQVKYDGDGDLIELSKKETENLRGVAKRCKSS 2212
QY 317 ----- 316
Db 2213 LYVVLVSVYCVMLASYANQSDSVGIPVSHRTHPQFSQVIGFFVNLVLRVDISQSAICG 2272
QY 317 -----MKSIMATQLRD----- 327
Db 2273 LIRRVMKELVDAQLHQMDFQFBVTKLQVDNDPSRHPLVQNVNPFESRANGEDHARSEDE 2332
QY 328 ----- 327
Db 2333 GSIAFNQYRPQVQVDSVAKFDLNAIVTELESGLRVNFNYATSLFNKSTIQGFLHTYEVLL 2392
QY 328 ----- 327
Db 2393 RQLSLSAEGINEDTQLSILVRPTENGDLHLPLAQSPPLATTABEQKVASINQAFEREAFLA 2452
QY 328 -----LATW- 331
Db 2453 AEKIAVQGDRLSYADLNGQANQLARYIOSVSGIGADGIALMLEKSIDTITICILATWK 2512
QY 332 ----- 331
Db 2513 AGAAVVLDPPTYPPGRVOLILBEEIKAKAVLVHSSHASKERHGAKVIAVDSPAIETAVSQ 2572
QY 332 ----- 331
Db 2573 QSAADLPTIASIGNLAYIIFTSGTSGKPGVLVEQKAVLLLRDLALRERYFGRCCTKHGV 2632
QY 332 ----- 331
Db 2633 LFLSNVYVDFSVQELVLSVLSGHKLIIVPPAEFVADDEFYRMASTHGLSYLSTGSPSLLOKI 2692
QY 332 -----VY----- 333
Db 2693 DLARLDHLQVVTAAGEELHATQYKMKRRRFPNGPIYNAYGVTTVTTVYNIIEAFTTNSIPEN 2752
QY 334 -----TTLRYRONPFCEP- 346
Db 2753 ALREVLPOTRAYLVNMAALQPVFPAVAGELYLAGDVTVRGYLNQPLLDQRFIPNPFCKEE 2812
QY 347 -----SR-----NRTAVSE----- 355
Db 2813 DIAMGRFARLYKTGDLVRSRFRNRQQQLLEVILGRGDLQIKMRGYRIEISEVQNVLTSSPG 2872
QY 356 -----FMKNTHVLI-----RNET----- 368


```

; : : : :
2873 VREGAVVAKYENNDTSRTAHSVGVYTTNETVSEADILTFMKARLPTVMVPSHLCCLE 2932
369 ---PYTIYGLDMSSL----- 381
2933 GALPVTINGKLDVRLPEINDSAQSSYSPRNIEAKMCLWESALGMERCGIDDDLFK 2992
382 ----- 381
2993 LGDSITSLHLVAQHNVGCKITVRDIFEHRTARALHDHVFMKDSRDNVTQFTEQGP 3052
382 -----YNETMFV----- 389
3053 VICEAPLLPIQDWFLSKALQHPMYNHTFYVTPPELDVDSLSAAVRDLOQYHDFMRKLK 3112
390 ----- 389
3113 REBVGVSQFAEDFSPAQLRVLVNKVDGSAANVEILDGWSGFNLGPIGSGYLHG 3172
390 -----ENKTASDSN 398
3173 EDRSARVWFSVHMAIDTVSQILVRDLQTLVRNGSLGSKGSSFWAEBAIQNYKASDSE 3232
399 KT----- 400
3233 RNHWKLVMETASSISALPTSTGSRVRLSRSLSPKTSALIOGGIDRQDVSVVDSLLTSV 3292
401 -----TPTSPM---GFQRTFD----- 415
3293 GLALQHIATGPSNVTIEGHGREVDQTLDSRTMGWFTTMYPPPEIPRLSTENIVQGVVA 3352
416 ----- 415
3353 VSRFRQVPARGVGYGLYGTQHPVQVTVNYLQGLARKQSKPKWVLAVGDNPEYGL 3412
416 ----- 415
3413 MTSPEKDRSSAVDTAVCTMIDVDSAWSLESEBQFISSEBGLNKILDGRASQ 3472
416 -----FLWDYLD----- 422
3473 TSEFPDVPQPAETTYTYFVELEPPROGPTLFLPPGEGGAESYFNIVKRLQRTNMVFN 3532
423 -----SLLFDEIRNFSRSP----- 438
3533 NYLHSHKRLRTPBELAEMVLDQVRGIPGHGYPHFICWSFGILAMEMSRRLVASDEKIGF 3592
439 -----TYVN-----LTPPEH 448
3593 LGIIDTYFNVRGATRTIGLDTEILDPIIH 3622

```

RESULT 12
US-08-222-617A-27
; Sequence 27, Application US/08222617A
; Patent No. 5882879
; GENERAL INFORMATION:
; APPLICANT: Veenstra, Annemarie E.
; APPLICANT: Martin, Juan P.
; APPLICANT: Garcia, Bruno D.
; APPLICANT: Gutierrez, Santiago
; APPLICANT: Barredo, Jose L.
; APPLICANT: Von Doehren, Hans
; APPLICANT: Palisea, Harriet
; APPLICANT: Van Liempt, Henk
; APPLICANT: Montenegro, Eduardo P.
; TITLE OF INVENTION: A Method for Influencing Beta-Lactam
; TITLE OF INVENTION: Antibiotic Production and for Isolation of Large
; TITLE OF INVENTION: Quantities of ACV Synthetase
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive

```

; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,617A
; FILING DATE: 04-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 97,157
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3727 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-222-617A-27

```

```

Query Match 17.1%; Score 420; DB 2; Length 3727;
Best Local Similarity 5.2%; Pred.No. 0.0084;
Matches 188; Conservative 68; Mismatches 163; Indels 3211; Gaps 53;

QY 8 VRDVPK-----MFVLISISF----- 22
DB 76 VNOVPERCDLSGLTDTSTRYQLASTGFGDASAAQERLMTVPVDVHAALQELCLERRSV 135
QY 23 ----- 22
DB 136 GSVINFSVHMLKFGNGTHTTITASLHREQLNQNSPWSVSPITVTHENROGWSVAQAV 195
QY 23 -----LLVSPIN-----CKV- 32
DB 196 ESIEAARGSEKSVTAIDSASSLVKMLPDLVSVFVDADDARIPCFDFPLAVIVRECDAN 255
QY 33 -----MSKALYN-----RPMRG-- 44
DB 256 LSLTLRFSCLFNEETICNFTDALNILLAEAVIGRVPVADIELLSEAEQKQLEENNTD 315
QY 45 ----- 44
DB 316 GEYPSKRLHLIEEVEHERHEDKIAVCDERELTYGELNAQNSLARYLRSIGILPEQLV 375
QY 45 -----LVLSKIGKYK-----LDQLK----- 59
DB 376 ALFLDKSEKLIIVTILGVKWSGAAYVPIDPTYDPERVRFVLDLDTKARAIIASNOHVERLQ 435
QY 60 -----LEILRQLETTISTKYNVSKQPVKNL-----TWNTFFPO--- 92
DB 436 EVIGDRNLCIIRLEPLASLAQDSKFPANLDDLPLTSQQLAYVYVTSSTGTCFPKGIK 495
QY 93 -----YYILAGP----- 99
DB 496 QHTNVNSITDLSARYGVAGQHEAILLPSACVPEFPVRQTLMALVNGHLLAVINDVEKY 555
QY 100 -----IQNYSIYVL-----WPDF----- 112
DB 556 DADTLLPPIRRHSITYLNGTASVLOQYDFSDCPSLNRIILVGENLTEARYLALRQPFKNR 615
QY 113 -----YSTQLRPAK-----YVYSQYNHTAKTITFRPPPCG 143
DB 616 ILNEVGFTESAFVTKLIFDPESTRKDTSLGRPVNRVKCYILN-----PSLK 662
QY 144 RVPSTMTCLSEM-----LNVSK----- 159
DB 663 RVP-IGATGELHIGGIGSKGYLNRLPELTPHPIPNPQFOTDCEKQIGNSLMVYKTDGLAR 721
QY 160 -----RNDTGEQCGNFTTFNPFNVPNRWTKL----- 188

```

Db 722 WLPNGEVEYLGRADFOIKLRGIRIEPGE-----IETMLAMYPRVRTSLVSKLRNG 773
QY 189 -----YVGPTK-----VND----- 198
Db 774 PEETTNEHLVGYVCDASVSADLLSFLKELPRYMIPTRLVQLSIPVNVNGKADURA 833
QY 199 -----SQTIFYFLG----- 206
Db 834 LPAVDISNTEVRSLRGTEIAGIEMADVLGARQSRVSRNDNFRGLGHSITCICLIA 893
QY 207 -----LTALLRYAQRNC-----THSFYL 225
Db 894 RIRQRQLSVSISVEDVFATRLERWADLLQNKQCKDKEAPTELEENAATDNYL 953
QY 226 VNAMSR----- 231
Db 954 ANSLQGGFYHYLKSMEQSDAYVMSQVLRNYNTLSPDLFORANKHAQSPPALRLRPSWE 1013
QY 232 -----NLFVY----- 236
Db 1014 KEVFOLLDDPPLDWRFLYFTDVAAGVEDRKLEDRRLRQDLTERFKLDVGRLFRVYLK 1073
QY 237 ----- 236
Db 1074 SENRFTCLFSCHAILDCMSLPLLEFKVHETYLQLLHGDNLTSMDDPYTRTORYLHAHR 1133
QY 237 -----PKY----- 239
Db 1134 EDHLDFAWGVQKINERCDMALLNERSRYKVQLADYDQEQBRHVHTIALSGDAWLADR 1193
QY 240 -----INGT----- 243
Db 1194 QCSAQGITLHSILOFWHVAHVHAYGGTHTITGTTISGRNLPILGIERAVGPYINTLPL 1253
QY 244 -----KLKNTMRKLKRKQAPVK----- 260
Db 1254 VLDHSTFKDKTIMEAIEDVQAKVNMVNRSGNVLGRHLKTKDLKHGLFDSLFLVLENYPNLD 1313
QY 261 ----- 260
Db 1314 KSRTLEHOTELGYSIEGGTEKLNYP LAVIAREVETGGFTVSYCIYASELFEEVMISELLH 1373
QY 261 -----EOFEKKAKK----- 269
Db 1374 MVQDTLMQVARGNEPVGSLEYLSSIOLEQLAAWNATEAFPTTTHEMFENERSQKPK 1433
QY 270 ----- 269
Db 1434 IAVVYBETSLTYRELNERANMAHQLRSDVSPNPNEVIALVMDKSEHMIWILAVWKSOG 1493
QY 270 ----- 269
Db 1494 AYVPIDPGYNDRIOYLEDTQALAVIADSCYLPRIKGMASGTLLYPSVLPANPDSKWS 1553
QY 270 -----TQSTTTPYFSYT----- 281
Db 1554 VSNPSPLSRSTDLAXIYIYTGTTGRPKGVTVHEHGVNVLQVLSKVFGLDRTDDEVILSF 1613
QY 282 -----TSAALNVTT-----NVTY----- 294
Db 1614 SNTVPDHFVEQMTDAILNGQTLVLDNGMRGDKERLYRIEKNRVITYLSTGLTPSVVSMVEF 1673
QY 295 -----SITT----- 298
Db 1674 SRFKDHLRRVDCVGEAFSPVPDKIRETFHGLVINGYGTVEYSITTHKGLYFPERRMDK 1733
QY 299 ----- 298
Db 1734 SIGQQVHNSTSVYLNEDMKRTPIGAVGELYLGEGVVRGYNRADVTAERFIPNPFQSEE 1793
QY 299 ----- 298
Db 1794 DKREGNSRLYKTDGLVRWIPGSSGEVEYLGRNDFQVKIRGURIEVGEIEAILSSYHGIK 1853

QY 299 -----AARR----- 302
Db 1854 QSVVIACDCREGAOKFLVGYVADAAALPSAAIRRFMQSRLPGYMWPSRLILVSKFPWTPS 1913
QY 303 ----- 302
Db 1914 GKLDTKALPABEESEIDVVPPRSEIERSLSCDIWASLELLEMPHEIIGYSDFFSLGDSLK 1973
QY 303 -----VSTSTI----- 308
Db 1974 STKLSFMIHESFNRAVSVALFCHRTVBEAQTHLILDAADVHEITPIDCNDTQMI PVSRA 2033
QY 309 -----AYRDPSSP----- 316
Db 2034 OERLLFIHFENGSNAYNIDAAPELPFGSVDAISLLEQALRGNLARHEALRTLVLKDHATGI 2093
QY 317 ----- 316
Db 2094 YLQKVLSPDEAQGMFSVNVDTAKQVERLDOETASLSQHVRLDDELPEARILKLESGL 2153
QY 317 ----- 316
Db 2154 YLILAFHHTCFDAWSLKVFEBELRALYALQKTKSAANLPALKAQYKEYALVHRQLSGD 2213
QY 317 ----- 316
Db 2214 RMENLSDFWLRLKILGLEPLQLITDRPRVQVKYDGDGDLSELSKKETENLRGVAKRKSS 2273
QY 317 ----- 316
Db 2274 LYVVLVSVVCMLASVANDSVSGIPVSHRTHPQFSQVIGFFVNLVLRVIDSQAICG 2333
QY 317 -----MKSIMATQLRD----- 327
Db 2334 LIRRVKMKELVDQLHQMDFOEVTKLQVNDPDSRHPPLVQNVFNFESRANGEHDAARSEDE 2393
QY 328 ----- 327
Db 2394 GSLAFNQYRPQVPDVSVAKFDLNATVTELSGLRVNFNYATSLFNKSTIQGFLHTYELL 2453
QY 328 ----- 327
Db 2454 ROLSELSAEGINEDTQLSLVRPTENGDLHLPLAOSPLATTABEQKVASINQAFREAFLA 2513
QY 328 -----LATW- 331
Db 2514 AEKIAVQGDRAISYADLNGOANQLARYIQSVSCIGADDDGIALMLEKSIDTIIICILAIWK 2573
QY 332 ----- 331
Db 2574 AGAAYVPLDTPYPPGVOLILEEIKAKAVLVHSSHASKCERHGAKVIAVDSPALETAVSQ 2633
QY 332 ----- 331
Db 2634 QSAADLPTIASLGNLAVIIFTSGTSGKPKGVLEQKAVLLRLDALRERYFGRDCTKHGV 2693
QY 332 ----- 331
Db 2694 LFLSNYVDFSVESQLVLSVLSGHKLI VPPAPAEFVADDFYRMASHTGHSYLSGTPSLLOKI 2753
QY 332 -----VY----- 333
Db 2754 DLARLDHLQVVTAAGEBLHATQEKMRFRFNGPIYNAYVTTTIVVNIIEAFTTNSIFEN 2813
QY 334 -----TTLRYRONPPCEP- 346
Db 2814 ALREVLPGTRAYVLTAALQVPFPAVAGELYLAGTVTRGYLNQPLLTQDQFIFNPFCKEE 2873
QY 347 -----SR-----NRTAVSE- 355
Db 2874 DIAMGRFARLYKTGDLVRSRFNRQQOQPOLEYLGRGDLOIKMRGVRIEISEVQNVLTSSPG 2933

```
QY 356 -----FMKNTVHL-----RNET----- 368
Db 2934 VREGAVVAKYENDTYSRTAHSLSVGYVTTDNETVSEADILTFMKARLPTVMVPSHLCCLE 2993
QY 369 ---PYTYGTLDMSL----- 381
Db 2994 GALPVTNGKLDVRRLEPEIINDSAQSSYSPRNIEAKMCRMLWESALGMERCGIDDDLFK 3053
QY 382 ----- 381
Db 3054 LGDSITSLVAQHNOVGCKITVRDIFEHRTARALHDHVFVKDSRDSNVTOFRTQGP 3113
QY 382 -----YYNETMFV----- 389
Db 3114 VIGEAFLPLIQDWFLSKALQHPMYNHTFYVRTPELDVDSAAVRDLOOYHDVFRMLK 3173
QY 390 ----- 389
Db 3174 REEVGFQSPAEDFSPAQLRVLVNKVDGSAVNEILDGWSGFNLNGPIGIGYLHG 3233
QY 390 -----ENKTASDSN 398
Db 3234 EDRSARVFWFVHMAIDTVSQILVRDLQTLYRNGSLGKSGSFRQWAEAIQNYKASDSE 3293
QY 399 KT----- 400
Db 3294 RNHWKLVMETASSISALPTSTGSRVRLSRSLSPKTSASLIQGGIDRQDVSVYDSLTSV 3353
QY 401 -----TPTSPSM-----GQRTFID----- 415
Db 3354 GLALQHIAPTGPSMVTIEGHRBEVDOTLDVSRMTGFWFTMYFPBIPRLSTENIVQVWA 3413
QY 416 ----- 415
Db 3414 VSRPRQVARGVGYGLYGTQHPQVTVNYLGLQARKQSKPKEMVLAVGDNEFEYGL 3473
QY 416 ----- 415
Db 3474 MTSPEKDRSSAVDVTAVCIDGTMIIDVDSAWSLESEQFISSIBGLNKILDGRASQ 3533
QY 416 -----PLMDYLD----- 422
Db 3534 TSRRPDVQPAETTYTFYFELEPRGQPTFLPPGEGGAEYFNNIVKRLRQTNMVF 3593
QY 423 -----SILFDEIRNFSRSP----- 438
Db 3594 NYLHKSRLTFTPELAEMYLQVRGIQPHGFHIGWSFGGILAMEMSRRLVASDEKIGF 3653
QY 439 -----TYVN-----LTPPEH 448
Db 3654 LGIIDTYFNVRGATRTIGLDTBILOPIH 3683
```

RESULT 13

```
US-08-876-991-2
; Sequence 2, Application US/08876991
; Patent No. 5925360
; GENERAL INFORMATION:
; APPLICANT: Gregor Meyers, Tillmann R menapf,
; APPLICANT: Heinz-J rgen Thiel
; TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,991
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,577
; FILING DATE:
; APPLICATION NUMBER: US/08/650,584
; FILING DATE:
; APPLICATION NUMBER: US/08/469,702
; FILING DATE:
; APPLICATION NUMBER: US/08/123,596
; FILING DATE:
; APPLICATION NUMBER: 07/797,554
; FILING DATE: 22-NOV-1991
; APPLICATION NUMBER: US 07/494,991
; FILING DATE: 16-MAR-1990
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: William M. Blackstone
; REGISTRATION NUMBER: 29,772
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-991-2
```

```
Query Match 17.1%; Score 420; DB 2; Length 3898;
Best Local Similarity 5.0%; Pred. No. 0.0094;
Matches 188; Conservative 87; Mismatches 175; Indels 3334; Gaps 58;

QY 2 GRKEM--MYRDVVK----- 13
Db 52 GRGDIRTLRLDLPKRCDSRGNHLPVSGIYIKPGVYQDYTGVPVYHAPLEFFDEAOF 111
QY 14 -----MFVLISIFLLV-----SFINC----- 30
Db 112 CEVTKRIGRVGSGKLYHYVCVDCILLKLAKRGTPRTLKWRNFTNCPLWVTSDD 171
QY 31 ----- 30
Db 172 GASGSKDKPDRMNKGLKIAPREHEKDSKTPPDATIVVEGVYQIKKKGVKGNQTD 231
QY 31 -----KVMKALY----- 38
Db 232 GLYHNKNKPPESRKKLEKALLAWAVITILLYQPVAAENITQWNLSDNGTNGIQRAMYLRG 291
QY 39 ----- 38
Db 292 VNRSLHGIWPEKICKGVPTHLATDTLKEIRGMDASERTNYTCCRLQRHEMHNKHCW 351
QY 39 -----NRP----- 41
Db 352 YNIDPMQLMNRQTQNLTEGPPDKECAVTCRYDKNTDVNVVTOARNRPTTLTGCKKGKNF 411
QY 42 -----WR 43
Db 412 SFAGTVIEGCPNFVNSVEDILYGDHECGSLLODTALYLLDGMTNTIENARQGARVTSWL 471
QY 44 GLVLS----- 48
Db 472 GRQLSTAGKGLERRSKTWFGAVALSPYCNVTRKIGYIWTNNCTPACLPKNTKIIGPKF 531
QY 49 ----- 48
Db 532 DTNAEDGKILHEMGHLSFEFLLSVLSDFAFETASTLYLILHYAIPOSHEEPGCDTN 591
```

QY 49 -----KIGKY-----KLQKLEILRQL- 66
Db 592 QLNLTVKLRTEDEVSSVMNIGKVCVRPDWPEYETKVALLFBEAQGVTKLVLRALRDLT 651
QY 67 -----FTT-----
Db 652 RVMSASTTAFCLIKVLGRQVVQGIWLLVTAQGRACKEDRYAISSTNEIGLLG 711
QY 70 -----
Db 712 AEGLTITWKEYSHGLQDDGTAKVACTAGSKFVTALNVVSRRYLASLHRALPTSVPTEL 771
QY 70 -----ISTKN-----
Db 772 LPDGTNPAIEMDDDFGLCPDTPSVIKGVKNTLLNGSAFYLVCPGVTGWVECTAV 831
QY 76 -----
Db 832 SPTTLRTEVVKTRDKPPHRVDCVTIIVEKEDLFHCKLGNWTCVKGDPVTVKGGQVK 891
QY 76 -----
Db 892 QCRWCFEKEPVGLPHYPHIGKILTNETGYRVVSDTCNRCGVISTEGEHECLIGNTT 951
QY 76 -----
Db 952 VKVHALDERLGMPCRPKEIVSSEGVKRTSCTFNKTKLRNKYPRDSYFQOYMLKE 1011
QY 76 -----
Db 1012 YQWFMNDVTDHHTDYFAEFVVLVVALLGGRVULWLVTVIITEQLAAGLOLQGEVV 1071
QY 76 -----VSKQPVKNLTWN-----
Db 1072 LIGNLITHDNEVVYFLLYLIVRDEPIKKWILLFHAMTNPNVKITVALLMISGVAK 1131
QY 88 -----TEF-----
Db 1132 GKIDGQWQRPVTSFDIQLALAVVVVVVLLAKRDPPTPLVTATLRTAKITNGFST 1191
QY 91 -----
Db 1192 DLVIATVSAALLTWTYISDYKXKTLQYLSTVTGIFLIRVLKIGELDLHAPTLPshr 1251
QY 91 POYIL-----AG-----PIONSYITLWF 110
Db 1252 PLFYLVLVISTAVTRWNLVDVAGLLQCVPTLLMVTWMADILTLILPTVELTKLY 1311
QY 111 -----DFYSTQLRKPARYV-SOYNHTAKTITFRP----- 139
Db 1312 LKEVKIGAERGWLKWTNKRVDIYEVDOQSEGVYLFPSKQRTSAITSLWPLIKAILIS 1371
QY 140 -----
Db 1372 CISKWQILVLLXLI FEVSYYLHKVIDEJAGTNFVSRVLAALIEVNWAFDNEEVKGLK 1431
QY 140 -----
Db 1432 KPELLSRVKELIHKHVRNEVVVWFGBEIEYGMFKLGLVKAATLSRNKHCMLCTVCE 1491
QY 140 -----PPCGR-----
Db 1492 DRWRGETCPKGRFGPPVCGMTLADPEKHKRIPIFEDQSGGPLREEHAGYLOVKAR 1551
QY 145 -----
Db 1552 GOLFLRNLPLATVKVQMLVGNLGTETIGLEHLGHWLRGPAVCKVKTBHERCTTSIMDKL 1611
QY 145 -----VPSMTCLSEM-----
Db 1612 TAPFGVMPRGTTFRAPVRPPTSLKIRRGLETGWAYTHQGGISSVDHVTGCKDLLVCDMT 1671

QY 156 -----NVSKRNDTGE-----QGC-----GNF 171
Db 1672 GRTRVWCOSNNKMTDESEYGVKTDGCGPEGARCVVFNPEAVNISGTKGAMVHLQKTGGEF 1731
QY 172 T-----TNPMPF-----NVPRWN-----TKLYVG-----PTKVNVDSTI-- 202
Db 1732 TCVTASGTAPFDLKNLKGWSGLPIPEASSGRVVGRVKVGKNEDESKPTKLMSIGIQTVS 1791
QY 203 -----
Db 1792 ATDLTEMVKITTNRGEFRQITLATGAGKTTELPRSVIEEIGRHRVLVLIPLRAAES 1851
QY 203 -----YFLGLTALLRYA-----
Db 1852 VYQVMRQKPSIAFNLRIGEMKGDMAITGITYASYGYFCOMSQPKLRAAWEVSPTFLDE 1911
QY 217 RNT-----HSFYLVNMSRN-----FRVPKYIN 241
Db 1912 YHCATPEOLAIMGKIHFR-----SENLRVAMTATPAGTIVTTGQKHPIEBEPIAEVNM 1965
QY 242 G-----TKLKNM-----RKLKR----- 255
Db 1966 GEDLGSYLDIAGLKIPVEEMKNMLVFVPTRNMAVEAAKLLAKAGYNSGYYSGEDPSN 2025
QY 256 -----QAP-----
Db 2026 LRVTQSQSVVVVATNAIESGVTLPDLDVVVDGLKCEKIRLSPKMPFIVTGLKRMV 2085
QY 259 VKEQFEKAK-----KTQST-----
Db 2086 IGEQAQRGRGRVGRVKPGYRYSQETPVGSKDYHYDLLQAQRYGIEDGINITKSFREMYD 2145
QY 274 -----
Db 2146 WSLYEEDSLMTQLEILNLLISELPMVAKNIMARTDHPBIQLAYNSYETQVPLFPK 2205
QY 274 -----TTPYFSY-----
Db 2206 IRNGEVTTIDYNYTFLNARKLGDDVPVYVATEDEDLAVELLGLDWDPPGQGTVEAGRA 2265
QY 282 -----
Db 2266 LKQVVLSTAENALLVALGYVGYQALSXKHI PVVTDIYSVEDHREDTTHLQVAPNAIK 2325
QY 282 -----TSAALNVTTNVT-----
Db 2326 TEGETELKELAQGDVORCVENVTNAREGQFMKSQLKVRTPYKETMTVADYVK 2385
QY 295 -----SITTAARV 303
Db 2386 FIEALTDKEDIKYGLWGAHTALYKSIGARLGHETAFATLVVWKLAFGGEISDHKQA 2445
QY 304 STSTIAY-----RPD-----SSFMKSIWATQRLDLATWYTTLYR----- 339
Db 2446 ATDLVYIINRQFPQDGTETOQEGRKFFVASLLVSA---LATYTKSWNNLNLKIVEPA 2502
QY 340 -----ONP 342
Db 2503 LATIPYAAKALKLPAPTRLESVVILSTAYKTYLSTIRGKSDGLLGTGVSAMEINSQNP 2562
QY 343 F-----CEPSNRNTASE-----FMQN-----THVLIRNETP----- 369
Db 2563 VSGIAMLGVGAHAHAHAIEASEQKRTLLMKVFNFLDQATDLBLV-ESPEKIIMAL 2621
QY 370 -----YTIYGTL-----
Db 2622 FEAVQTVGNPLRVLYHYGVYKVGWEAKELAQRTAGRNFTLIMFEAVELLGVDSEKIR 2681
QY 377 DMSS-----
Db 2682 QLSNNILELLYKFRDNKIKSVREIAISWAPFSCDWTPTDDRIGLPHDNLVRYETKCP 2741
QY 381 ----- 380

```

Db 2742 CGYRMKAVKNCAGELRLLEBGGSPLCRNKFGSQNRVTKYDDNLSEIKPVIRMEGHV 2801
Qy 381 -LYY----- 383
Db 2802 ELYYKGATIKLDFNNSTKVLATDKWEYDHSYLRALAKRYTGAGYRGAYLGEKPNHKLIIQ 2861
Qy 384 ----- 390
Db 2862 RDCATITKDKVCFTKMRGCAFTYDLHLNLRLIELVHKNLEDRIPAVTWTMLAYT 2921
Qy 391 ----- 404
Db 2922 FVNEDIGTIKPTGKVTPEKBEVLQPAVVDTTDVAVTVVGTSTMTTGETPTFTS 2981
Qy 405 ----- 414
Db 2982 LGSDSKVRQVLKLGVDGQYFPGPNQQRASLLEALQGVDERPSVLLGSDKATSNRVKTAK 3041
Qy 415 ----- 438
Db 3042 NVKIYRSRDPLELRMMKRGKILVVALSRVDLTALLKEVDYKGTFLTRTLEALSGLKPKK 3101
Qy 439 ----- 438
Db 3102 RDITKAEQWLLRLEDOIELBPDWFAKSPIFLEANTIKRDYHLVGDIAITIKAKAQOLGA 3161
Qy 439 ----- 438
Db 3162 TDSTKISKEVGAKVYSMKLSNWVIOBENKOGSLAPLPEELLQCPGQGNKTHMVSAYQ 3221
Qy 439 ----- 443
Db 3222 LAQGNWVPVSVCHVMGTIPARRTKTHPEYAVKLRELVDHMKALCGGSLGSKHNEWI 3281
Qy 444 ----- 443
Db 3282 GKVKYQGNLRTKHLNPKGVAQLHREGYRHNVNKTIIGSVMTATGIRLEKLPVVRQTD 3341
Qy 444 ----- 450
Db 3342 TTNFHOAIRDKIDKEENLQTPGLHKLMEVFNALKRPELEASDVANDWELBERGINRKA 3401
Qy 451 ----- 450
Db 3402 AGFFERKNIGVLDSEKKNVEEVIDSLKGNRIYYETAIPKNEKRDVNDWTAGDFVDE 3461
Qy 451 ----- 450
Db 3462 KKPRVIQYPEAKTRLAITKVMYKWKQKPVVIFGYEGKTPFQIFDKVKKEWQFQNPVA 3521
Qy 451 ----- 450
Db 3522 VSFTKAWDTQVTRDLELRDIQKPYFKKKWHKPIDTLTKHSEVPVISADGEVIRKG 3581
Qy 451 ----- 450
Db 3582 QRGSGQPDTSAGNSMLNVLTMVYAFCEATGPYKSPDRVAKIHVCGDDGFLITERALGEK 3641
Qy 451 ----- 450
Db 3642 FASKGVQILYEAGKPKQKITEGDKMKVAYQDDIEFCSHTFPVQVRWSDNTSSYMPGRNTTT 3701
Qy 451 ----- 450
Db 3702 ILAKMATRLDSSGERGTIAYEKAVAFSFLMYSNWPNLIRICLLVLSTELQVPRGKSTTY 3761
Qy 451 ----- 457
Db 3762 YBEGDPISAKEYVIGHNLFDLKRTSPFKLAKNLMSLTGLVWTRHTSKRLLQDCVNVGTK 3821
Qy 458 NSLW 461

```

```

Db 3822 EGNW 3825

RESULT 14
US-09-059-853-2
; Sequence 2, Application US/09059853
; Patent No. 5935582
; GENERAL INFORMATION:
; APPLICANT: Gregor Meyers, Tillmann R menapf,
; APPLICANT: Heinz-J igen Thiel
; TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Organon Teknika Corporation
; ADDRESSEE: Biotechnology Research Institute
; STREET: 1330-A Piccard Drive
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,853
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: 07/797,554
; APPLICATION NUMBER:
; FILING DATE: 22-NOV-1991
; APPLICATION NUMBER: US 07/494,991
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: William M. Blackstone
; REGISTRATION NUMBER: 29,772
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-059-853-2

Query Match 17.1%; Score 420; DB 2; Length 3898;
Best Local Similarity 5.0%; Pred. No. 0.0094;
Matches 188; Conservative 87; Mismatches 175; Indels 3334; Gaps 58;

Qy 2 GRKEM--WVRDVPK----- 13
Db 52 GRGDIHTLRDLPRKGDGCRSGNHLGVPVGIYIKPGPVYQDYTGVPVYHRAPLPFFDEAQF 111
Qy 14 ----- 30
Db 112 CEVTKRIGRTGSDGKLYHYVCVDCGICLLKLAAGTPTLKWIRNFTNCPPLWVTSDD 171
Qy 31 ----- 30
Db 172 GASGSKDKKPDNRNKGKLIAPREHEKDSKTKPPDATIVVEGVKYQIKKKGVKGNKNTQD 231
Qy 31 ----- 38
Db 232 GLYHNKPKPEPRKKLEKALLAWAVITILLYQPVAAENITQWNLSDNGTNGIQRAMVLRG 291
Qy 39 ----- 38
Db 292 VNRSLLHGIWPEKICKGVPTHLATDTLKEIRGNMDASERTNYTCCRQRHEWNGHCNW 351
Qy 39 ----- 41

```

Db 352 YNIDPWIQMNRQTNLTEGPPDKCAVTCRYDKNDVNVVTOARNRPTTLTGCKKGKF 411
QY 42 -----
Db 412 SFAGTVIEGPCNFVNSVEDIYDHECGSLLODTALYLLDGMWNTNIENARQARVTSWL 471
QY 44 GLVLIS-----
Db 472 GRQLSTAGKLRERRSKTWFGAVALSPYCNVTRKIGYIWTYNNCTPACLPKNTKIIGBKF 531
QY 49 -----
Db 532 DTNAEDGKILHEMGHLSFLLSLVILSDPAPETASTLYLLHYALPQSHERPBGCDTN 591
QY 49 -----KIGKY-----
Db 592 QLNLTVKLRTEDVFPSSVWNLGKVCVRPDWVPYETKVALLFBEAGQVILVLRALDLT 651
QY 67 -----ETT-----
Db 652 RVWNSASTTAPFLICLIKVRGOVVQGIIMLLLVTAQGLRACKEDYRYAISSTNEIGLLG 711
QY 70 -----
Db 712 AEGLTWTKEYSHGLQLDDGTAVXAVCTAGSKFVKTALNVVSRRLASLHKRALPTSVTFEL 771
QY 70 -----ISTKYN-----
Db 772 LFDGTPAIBEMDDDFGLCLPDPSTPVIKGKNTLLNGSAFVLVCPIGTWVVECTAV 831
QY 76 -----
Db 832 SPITLRETVVKTFRDKPPHRVDCVTTIWEKEDLFHCKLGNMTCVKGDPTVYKGQVK 891
QY 76 -----
Db 892 QCKWGFEPKPYGLPHYPHIGKILNETGYRVVSDTCNDRDGVVISTEGEHECLIGNTT 951
QY 76 -----
Db 952 VKVHALDERLGPMPCKPEIVSSEGPVKTSCTFNKYTKLRNKYEPDSYFOQYMLKGE 1011
QY 76 -----
Db 1012 YOYWFNLVTDHDTDYFAEFVVLVVALLGGYVVLVITVYIILTEQLAAGLQLOGEVV 1071
QY 76 -----VSKOPVKNLTWN-----
Db 1072 LIGNLITHDNEVVYFLLLVIRDEPIKKWILLPHAMTNPNPKTITVALLMISGVAK 1131
QY 88 -----TEP-----
Db 1132 GKKIDGGWQROPVTSFDIQLALAVVVVVMLAKRDPPTFPLVITVATLRTAKITNGFST 1191
QY 91 -----
Db 1192 DLVIATVSAALLTWYISDYKKTWLOYLNVSTVGTGIFLIRVLKGIGELDLHAPLPSHR 1251
QY 91 POYVIL-----AG-----
Db 1252 PLFVILVYLSTAVTRWNLVAGLLQCVPTLLMVFTWADILTLILPTFYLTKLY 1311
QY 111 -----DFYSTQLRKAKVYV-SQYNHTAKTITFRP-----
Db 1312 LKEVKIGAERGLWKTNYKRVNDIYEVDQTSBGVYLFPSKORTSAITSTMPLIKAILIS 1371
QY 140 -----
Db 1372 CISKWKQIYLLYLIFEVSYLHKVVIDEAGTNFVSRVLAALIEVNPADNEEVKGLK 1431
QY 140 -----

Db 1432 KPFLSSRRVKELIHKHVRNEVVVWVWFGEDEIYGMPLKILGLVKAATLSRNKHCMLCTVCE 1491
QY 140 -----PPCGR-----
Db 1492 DRDMRGETCPKGRFGPPVCGMTLADPEEKHYKRIPIREDQSGGPLREHAGYQVKAR 1551
QY 145 -----
Db 1552 GOLFLRNLFLVATKVKMLLVGNLGTGIDLEHLGWLGRPAVCCKVTEHERCTTSMIDL 1611
QY 145 -----
Db 1612 TAFGVMPRGTTTPAPVRFPFPTSLIKTRGLETCWAYTHQGISVSDVHTVTCGKDLLVCDTM 1671
QY 156 -----NVSKRNDTGE-----QGC-----
Db 1672 GRTRVQCNSKNMTDSEYGVKTDSCGPEGARCYVFNPEAVNISGTGKMWHLQKTGGEF 1731
QY 172 T-----TFNPMFF--NVPWN-----TKLYVG-----PTKVNVDSTI--- 202
Db 1732 TCVTASGTPAFDLKWLKWSGLPIFEASSGRVGRVKVGNEDSKPTKLMSGIQTVS 1791
QY 203 -----
Db 1792 ATDLTEMVKITTMNRGEPQITLATGAGKTTELPRSVIEISGRHKRVLVLIPLRAAES 1851
QY 203 -----YPLGLTALLRYA-----Q 216
Db 1852 VYQYMRQHPSTAFNLRIKEMKEGDMATGITYASYGYFCQMSQPKLRAAMVVEYSFELDE 1911
QY 217 RNC-----HSPYLVNAMSRL-----SENLRVAMTATPAGITVTTTQXHPIEBTAPEVMK 1965
Db 1912 YHCATPEQLAIMGKTHRF-----
QY 242 G-----TKLKNTM-----RKLKRX-----
Db 1966 GEDLSEYLDIAGLKIPEVEMKNNLVFVPTNRNNAVEAAKLRKAGVNSGYYSGEDPSN 2025
QY 256 -----QAP-----
Db 2026 LRVVTSQSPVYVATNAIESGVTLPLDLDVVVDTGLKCEKRILSPKMPFIVTGLKRMVAT 2085
QY 259 VKEQFEKKAK-----KTOST-----
Db 2086 IGEQAQRGRVGRVKPYRYSOETPVGSKDHYDLLOAQRYIEDGINITKSFREMYD 2145
QY 274 -----
Db 2146 WSLYEEDSLMITOLEILNLLISEELPMVAKNIMARTDHPPIQLAYNSYETQVPVLPFK 2205
QY 274 -----TTPYFSYT-----
Db 2206 IRNGEVTDYDNTFLNARKLGDVFPYVYATEDDEDLAVELLGLDWDPCGQGTVEAGRA 2265
QY 282 -----
Db 2266 LKQVGLSTAENALLVALFGYVGYQALSKRHIPVVTDIYSVEDHRLEDTHLQYAPNAIK 2325
QY 282 -----TSAALNVTTNTVY-----
Db 2326 TEGKETELKLAQGVQRCVEAVNYAREGIFQMKSQLKVRTPTYKETMNTVADVVK 2385
QY 295 -----SITTAARRV 303
Db 2386 FIEALDTSKEDIKYGLWGAHTALYKISGARLGHETAPATLVVKNLAFGESISDHKQA 2445
QY 304 STSTIAY-----RPD-----SSPMKSIATQRLDRLATWYTYTLRY----- 339
Db 2446 ATDLVWYIINRPOPFGDTQOQGRKFVASLLVSA---LATYTKSWNTNLSKIVEPA 2502
QY 340 -----QNP 342
Db 2503 LATLPYAAKALFLFAPTRLESVILSTAIYKTVLSTIRGKSDGLLGTGVSAAEIMSQNP 2562

343 F-----CEPSNRRTAVSE-----PMKN-----THVLIRNETP-----369
 2563 VSVGIAVMLGVGAAVAAHNAIEASEQKRTLLMKVFNFLDQAATDELVK-ESKEIIMAL 2621
 370 -----YIYIGTL-----376
 2622 FEAVQTGNNPLRLVYHLGVFYKWEAKELAQRTAGRNLFLLIMFEAVELLGVDSSEKIR 2681
 377 DMS-----380
 2682 QLSNLYLELLYKFRDNKSSVREIAISWAPAFSCDWTPTDDRIGLPHDNYLRVETKCP 2741
 381 -----380
 2742 CGYRKAVKNCAGELRLLEGGSFCLCNKFRGSGNRYVTYYDDNLSEIKPKVIRMEGHV 2801
 381 -LY-----383
 2802 ELYYKGAATIKLDPNNSKTVLATDKWEVDHSTLVRALKRYTGAGYGLGEKPNHKLHIQ 2861
 384 -----NETMFE-----390
 2862 RDCATITKDKVCFIKMKRGCAFTYDLSLNLTRLIELVHKNNLEDRIPAVTWTTLAT 2921
 391 -----NKTASDSNKTPTTS-----404
 2922 FVNEDIGTIKPTGKVTPEKQEVVLQPAVVVDDTDVAVTVVGESTWTGTPTTTS 2981
 405 -----PSMGFORTFI-----414
 2982 LGSDSKVRQVLKLGVDGQYGPNGQASLLEAIOGVDERPSVLILGSKATSNRVKTAK 3041
 415 -----DPL-----WDVDSLLFLDEIRNFSRSP--438
 3042 NVKIYRSRDLREMMKRGKILVVALSRVDTALLKFDYKGTFLTRTEALSGLKPKK 3101
 439 -----438
 3102 RDITKABQWLLRLDQIBELPDWPAKEPIPLEANIKRDKYHLVGDVATIKERAKQLGA 3161
 439 -----438
 3162 TDSTKISKEGAKVYSKMLSNWVIOBENKQSLAPLFEELLOQCPGQNKTHMVSAYQ 3221
 439 -----TYVNL-----443
 3222 LAQGNVVPVSVCHVFMGTIPARTKTHPYEAYVKLRELVDHMKKALCGSGLSKHNWVI 3281
 444 -----443
 3282 GKVKYQGNLRTKHLNPGKVAEQHLHREGYRHNVYNTIGSVMTATGIRLEKLPVVRQTD 3341
 444 -----TPPEHRR-----450
 3342 TTNFHOAIRDKIDKEENLQTPGLHKKLMEVFNALKRPELEASDYDAVDWHEELERNGKA 3401
 451 -----450
 3402 AGFFERKNIGEVLDSKKNVEVDSLKKGRNIRYETAIIPKNEKRDVNDWTAGDFVDE 3461
 451 -----450
 3462 KKPRVIQYPEAKTRLAITVMYKWKQKPVVPIPGYEGKTPLFQIPDKVKKENDQFQNPVA 3521
 451 -----450
 3522 VSPDKAWDTQVTTROLELRIDQKVFYKKWKHKFIDTLTKHMEVSPVTSADGEVYIRKG 3581
 451 -----450
 3582 QRGSGQPDTSAGNSMLNVLTMVYAFCEATGVPYKSFDRVAKIHVCGDGDGFLITERALGEK 3641

451 -----450
 3642 FASKGVQILYEAGKPOKITEGDKMKVAYQFDDIEFGCHSTPQVVRWSDNTSSYMPGRNTTT 3701
 451 -----450
 3702 ILAKMATRLDSSGERGTIAYEKAVAFSFLMYSNWNLPIRRICLLVLSTELQVRPGKSTTY 3761
 451 -----AVNLSTS 457
 3762 YVEGDPISAYKEVIGHNLFDLKRTSPKLAKNLSMSTLGLVWTRHTSKRLLQDCVNVGTK 3821
 458 NSLW 461
 3822 EGNW 3825

RESULT 15

US-08-476-515A-84
 ; Sequence 84, Application US/08476515A
 ; Patent No. 6239270
 ; GENERAL INFORMATION:
 ; APPLICANT: Akerstrom, Goran
 ; APPLICANT: Jublin, Claes
 ; APPLICANT: Raab, Lars
 ; APPLICANT: Crumley, Gregg R.
 ; APPLICANT: Morse, Clarence C.
 ; APPLICANT: Murray, Edward M.
 ; APPLICANT: Hjaln, Goran
 ; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
 ; TITLE OF INVENTION: Thereof and DNA Encoding Same
 ; NUMBER OF SEQUENCES: 84
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Martin Savitzky
 ; STREET: Rhone-Poulenc Rorer Inc., 500 Arcola Rd.;
 ; STREET: 3C43,
 ; CITY: Collegeville
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19426-0107
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Compaq PC
 ; OPERATING SYSTEM: Windows 95
 ; SOFTWARE: Word 7.0 (Patentin)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/476,515A
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/344,836
 ; FILING DATE: 23-NOV-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: WO PCT/SE94/00483
 ; FILING DATE: 24-MAY-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: SE 9301764-8
 ; FILING DATE: 24-MAY-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Savitzky, Martin
 ; REGISTRATION NUMBER: 29,699
 ; REFERENCE/DOCKET NUMBER: A1355D
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 610-454-3816
 ; TELEFAX: 610-454-3808
 ; INFORMATION FOR SEQ ID NO: 84:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4654 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-476-515A-84

```
Query Match      17.0%; Score 417; DB 3; Length 4654;
Best Local Similarity 4.3%; Pred. No. 0.019;
Matches 191; Conservative 75; Mismatches 152; Indels 4068; Gaps 63;

QY 8 VRD----- 10
Db |||
130 VRDPCGADNDCCQPTCEQLTCDNGACVNTSQKCDKWKVDCDSDEINCTEILCHNEFS 189
QY 11 -----VPMKFV-----
Db : : :
190 CGNGECIPRAYVCDHNDCCQDSDEHACNYPTCGGYQFTCPSGRCIYNWVWCGEDDCKD 249
QY 17 -----LISI-----SF 22
Db |||
250 NGDEDCESGPHDVHKSPREWSCEPSCGRCISIVKVCGLDCPGRDENNTSTGKYSM 309
QY 23 LUVSPINCK-----VMSKALYNRP--- 41
Db : : :
310 TLCNALNCQYCHETPYGACFCPPGYIINHDSRTCVEPDDCQIWIQICDQKCESRGRH 369
QY 42 ----- 41
Db 370 LCHCEGYLBERGOYKANDSFGBASIFNSGRDLLIGDIHGSRFILLVESQNRGVAVGV 429
QY 42 -----WRGLVLSKI----- 50
Db 430 APHYLQRFVETDQVQKFSVDINGLNTQEVNLVSVETPENLAVDVWNNKIYLVETKVN 489
QY 51 -----GKVKL----- 55
Db 490 RIDMNLDSYRVTLITENLGHPRGIAVDPTVGYLFFSDWESLGEPLERAFMDGNSRK 549
QY 56 ----- 55
Db 550 DLVKTCLGWPAGVTLDMISKRVYVWDSRPDIETVTVYDGIQRTVHGGSLIHPFGVSL 609
QY 56 -----DOLKLEIL----- 64
Db 610 FEGQVFTDWTQWVLKANKFTETNPQVYQASLRPGYVTVYHSLRQPYATNCKNNGG 669
QY 65 -----QLET-----TIST- 72
Db |||
670 CEQVCVLSHRTDNLGFRCKTGFQDLTDTERHICIAVQNLIFSSQVAIRGIPFTLSQ 729
QY 73 -----KYNVSKQPV----- 81
Db 730 EDVMVPSGNSPFFVGIDFDAQDSTIFFSDMSKMFQKIDGTGREILAAKNRVENVESL 789
QY 82 -----KMLTNTEFPQY-----ILAGPIQNYSI 105
Db 790 AFDWISKNL-----YMTDSHYKISVMRLADKTRRTVVOYLNPNRSVVVHPGAYLF 841
QY 106 TYLWFDYFVQLRKPQYV--YGOYNHTAKTI-----
Db |||
842 FTDWF-----RPAKIMRWSGDGSHLLFVINTLIGWPNGLAIDWAASRLYWDAYFDK 893
QY 136 ----- 135
Db 894 IEHSTFDGLRRRLGHIEQWTHPGLAIFGEHLFFTDWRLGAIIRVRKADGEMTVIRSG 953
QY 136 ----- 135
Db 954 IAYILHLSYDVNIQTGSNACQPHTPNGDCSHFCFPVFNQFVCGCPYGMRLASNHLTC 1013
QY 136 -----PC-GR-VPSM-----TCLSEML- 155
Db 1014 EGDPTNEPTEQCGLFSFPCKNRCVNPYLLCDGVDCHDNDSEOLCGTLNNTCSSAFT 1073
QY 156 -----NVSKRND----- 162
Db |||
1074 CGHGECIPAHWRCKENDVDGDEHNCPHAPASCLDTQYTCNHNQCSKNWVCDTND 1133
QY 163 -----TGEQCGCNFTTFNMPFNVR----- 183
```

```
Db 1134 CGDGDSEKNCNSTETCQPSQFNCNHRICIDLFSVCDGDKDCVDGSDVEGCVLNCASQPK 1193
QY 184 ----- 183
Db 1194 CASGDKCIGVTNRCDGVFDCSDNSDEAGCPTRPPGMCHSDEFQCEQDGGICIPNFWECDGH 1253
QY 184 ----- 183
Db 1254 PDCLYSGDEHNACVPKTCPSYFHCNNGNCIHRAMLCDRNDNDGDMSEKDCPTQTPRCP 1313
QY 184 -WNTK----- 187
Db 1314 SWQOCGLGHNICVNLVSVCDGIFDCPNGTDESPLCNGNSCSDFNGCTHECVQEPFGAK 1373
QY 188 ----- 187
Db 1374 LCPLGFLANDSKTCEDIDECILGSCQHCYNMRGSRFCSDTGYMLESDCRTCKVTAS 1433
QY 188 -----LYVGTQKVVDSQT-----IYFL----- 205
Db 1434 ESLLLLVASQNKIIADSVTSQVHNISLVENGSIYAVDSDISGRIFWSDATOGKTWSA 1493
QY 206 ----- 205
Db 1494 FQNGTDRRVVFDSSIIILTETIAIDWVGRNLYTWTDYALETIEVSKIDGSHRTVLISKNLN 1553
QY 206 -----GLTA----- 209
Db 1554 PRGLALDPMEHLLFMSDGHHPRIERASMDGSMRTVIQVDKI FWPCLGJTIDYFNRLLY 1613
QY 210 -----LLRY----- 214
Db 1614 FMSYLDYMFCDYNGHHRQVIAASDLIIRHPALFLFEDSVYTTDRTRVRANKWHG 1673
QY 215 -----AQRNCTH-----SFY----- 224
Db 1674 GNQSVWYNTQWPLGIVAVHPKQPNVSNPCAFSCSHLCLLSQGHFYSVCPSGWSL 1733
QY 225 -----LVNM----- 229
Db 1734 SPOLLNCLRDOPFLITVRQHIIFGILNPEVKSNDAMVPIAGIQNGLDVDFDDEQYIY 1793
QY 230 -----SRNLFR----- 235
Db 1794 WVENPGEIHRVKTGTNRNRTVFASISWVGPSSMLALDWISRLNLYSTNPTQTSIEVLTJLHG 1853
QY 236 -----VPKYI-----NQTJLK-- 246
Db 1854 IRYRKTLIANDGTALGVGPPIGIVDPARGKLYNSDQDTSQGVPAKIASANMDGTSVKTL 1913
QY 247 ----- 246
Db 1914 FTGNLEHLECVTLIDIEQKLYWAVTGRGVIERGNVDGTRMILVHLSHPWGLAVHDSFL 1973
QY 247 ----- 246
Db 1974 YVTDEQVEVIERVDKATANKIVLRDWNVNLRLQVYHRRNAEASSNGCSNNNACQQIC 2033
QY 247 ----- 246
Db 2034 LPVPGFLFSCACATGPKLNDPDRSCPYNSFIVVMSLAIRGFSLELSHSETMVPVAGQ 2093
QY 247 -----NTMRKLK----- 253
Db 2094 GRNALHVDVDSGFIYWCDFSSVASDNAIRRIKPDGSSLMNIVTHGIGENGVRGIAVD 2153
QY 254 -----RK-----QAP 258
Db 2154 WVAGNLYFTNAFVSETLIEVLTINTYRVLKVTVDMPRIHVDPKRYLFWADYQRP 2213
QY 259 VKE----- 261
```


Db 2214 KIERSPDCTNRTVTVSEGIPTVPRGLAVDRSGVYVWVDDSLDIARIRINGENSEVIRY 2273
QY 262 -----QPEKKAKTQSTT----- 274
Db 2274 GSRVPTPGITVPENSIIWDRNLKKIIFOASKEPENTPEPTVIRDININWLDRVTIPDKQV 2333
QY 275 -----TP----- 276
Db 2334 QPRSPAENVNNPCLENNGGSHLCFALPGIHTPKDCACFTLOSOGKNAISTENFLIPA 2393
QY 277 -----YF-----SYTT-SA 284
Db 2394 LNSLSRLHLDPENHSPPFQTIINVERTVMSLDYSVSDRIYFTQNLASGVQISYATLS 2453
QY 285 ALNVTNVTYSITTA-----ARRV----- 303
Db 2454 GIHTPTVIASIGITADGIAFDWITRRIYSDYLNQMINSMAGDSNRTVIARVPKPAIRV 2513
QY 304 -----STSTIARPDSSPMKSI MAT 323
Db 2514 LDPCCQGLYWADWDTHAKIERATLGGNFRVPIVNSSLVMPSGTL DYE----- 2561
QY 324 QLRDLATW-----YTTLAVRON 341
Db 2562 --EDLLYWVDASLORIERSTLTGVDRREVIVAAVHAFGLTYGYIYWTDLTYTORIYRAN 2619
QY 342 PF----- 343
Db 2620 KYDGGQIAWTTNLLSQPRGINTVVKNOQOCNCPCEQFNGGCSHICAPGNGAECQCPH 2679
QY 344 ----- 343
Db 2680 EGNWYLANNRKHCIVDNGERGASFTCSNGRCISEWKCDNDNDCGDSDEMESYCALH 2739
QY 344 -CEPS-----RNR TAVSEFMKN----- 359
Db 2740 TCSPTAFTCANGRCVQYSYRCDYNDGCGSDEAGCLFRDCNATTEFCMNRRCIPREFI 2799
QY 360 ----- 359
Db 2800 CNGVDNCHDNTSDEKNCPRDTCQSGYTKCHNSNICIPRYLDCDNDNDCGDSNDENPTYC 2859
QY 360 -THVLTRNE----- 367
Db 2860 THTCSSEFQACASGRICPOHWYCDQETDCFDASDEPASCGRHSERTCLADEFKDGGRCI 2919
QY 368 ----- 367
Db 2920 PSEWICDGNDCGDSDEDKRHQCQONQCSDBFLCVNDRPPDRRCIPQSWCVDGDV DCT 2979
QY 368 ----- 367
Db 2980 DGYDENQNCNTRTCSNEFTCGYGLCIPKIFRCDRHNDGDSYDERGCLYQTCQONQFTC 3039
QY 368 ----- 367
Db 3040 QNGRCISKTFVCDNDCGDSDELMLHLCHTPEPTCPPHEFKDNGRCIEMMKCNHLDD 3099
QY 368 ----- 367
Db 3100 CLDNDSEKGGINECHDPSISGCDHNCTDTLTSFYCSRPGYKLSMDKRTCDVIDBCTEM 3159
QY 368 -----TPYTIYGLDMSLYY----- 383
Db 3160 PFVCSQKCVENIGSYICKCAPYLREPDGKTCRQNSNIEPYLIP-----SNRYLRNLTI 3214
QY 384 -----NETMFVNKT----- 393
Db 3215 DGYFYSILLEGLDNVALDPDRVEKRLYWDTORQVIERMFL-NKTNKETIINHLRPAAE 3273
QY 394 ----- 393
Db 3274 SLAVDWVSRKLYWLDARLDGLFVSDLNGGHRRLAQHCVDANNPTFCFDPNRLGLALHPQYG 3333

QY 394 ----- 393
Db 3334 YLYWADGHRAYIGRVGMDGNTKSVIISTKLEWPGITTIIDYTNLLYWAHAHLGYIEYSD 3393
QY 394 ----- 393
Db 3394 LEGHHRHTVYDGAHPHPFAITIPEDTIYWTDMNTRTVKGNKYDGNRQTLVNTTHRPFD 3453
QY 394 ----- 393
Db 3454 IHVYHYPQPIVSNPCGTNGGCSHLCLIKPGKGFTCECPDDFRTLQLSGSYCYMPCMS 3513
QY 394 -----ASDSNKTTT----- 402
Db 3514 STOFLLCANNEKCIPIWVKCDGKDCSDGDELALCPQRFCLRQFQCSGDNCTSPQTLCN 3573
QY 403 ----- 402
Db 3574 AHQNCPDGSEDRLLCENHHCSNEWQCANRCIPESWQCDTTPNDCEDNDEDESSHCASR 3633
QY 403 ----- 402
Db 3634 TCRPGQFRANGRCIPQAMKCDVDNDCGSDHSDPEIECHSSAHLCDNFTFSCKNYRCI 3693
QY 403 ----- 402
Db 3694 PKWAVCNGVDDCRDNDSEOGCEBERTCHPVGDPRCKNHHCIPLRWQCDGQDNDGNSDEN 3753
QY 403 ----- 402
Db 3754 CAPRECTESEFRVNOQCIPSRWICDHYNDCGNSDERCEMRTCHPEYFQCTSGHCVHS 3813
QY 403 ----- 402
Db 3814 ELKCDGSADCLDASDEADCTRPFGAYCOATMFECKNHVCIPPYWKCDGDDCGDSDE 3873
QY 403 -----TSP----- 405
Db 3874 ELHLCLDPCNSPFRCDNNRCIYSHVCNGVDCGDTDETEHCKRPTPKPCTEY 3933
QY 406 ----- 407
Db 3934 KCGNGHCIPHNDVDDADCGDMSDELGCNKGKERTCAENICEQNTQLNEGGFICSTA 3993
QY 408 GFORTFID----- 415
Db 3994 GFTNVFDRITSLDINECEQFGTCPOHCRNTKGSYECVADGFTSMDSRPGKCAEGSS 4053
QY 416 PL----- 417
Db 4054 PLLLLPDNVIRKYNLSSRFSYQLQDEBYIOAVDYDWDPKDIGLSVYVTVRGEGRFG 4113
QY 418 ----- 417
Db 4114 AIKRAYIPNFESGRNNLVOEVDLKLKYVMQPGIAVDWGRHYIYWSVKNKRIEVAKLGD 4173
QY 418 -----W----- 418
Db 4174 RYRKMLISTDLQPAIAVNPGLMLFWTWDGKPKIESAWMNGEDRNILVFDLGNPTG 4233
QY 419 ----DYLDLSLF----- 426
Db 4234 LSIDYLDRIYWSDFKEDVIETIKYDGTDRRVIAKEAMNPYSLDIPEDQLYMSKEGEV 4293
QY 427 -----LDEIRNFS-----LRSTY----- 440
Db 4294 WKQNFQGGKSKTLVWNPMLTQVRIFHQLRYNKSVPNLCKQICSHLCULLRPGGSCACP 4353
QY 441 -----VNLTP----- 446
Db 4354 QSSSFIEGSTTECDAAIELPILPPPCRCMHGNCYFDETDLPKCKCPSGYTKYCEMAF 4413

Qy	447	-----EHRRAVNL-----	454
Db	4414	SKGISPGTTAVAVLLTILLVIVIGALAIAGFFHYRTGSLLPALPKLPSSLVKPSENG	4473
Qy	455	-----	454
Db	4474	NGVFRSGADLNMIDIGVSGFPGFETAIDRSMAISEDVVMEMGKQPIIFENPMYSARDSAVK	4533
Qy	455	-----	
Db	4534	VVQPIQVTVSENVNDKNYGPINPSEIVPETNPTSPAADGTQVTKW	4579

Search completed: December 7, 2004, 15:41:38
Job time : 131 secs